Specification and estimation of structural econometric models of the labour market

Bloemen, Hans Gerald

Publication date:
1994

Document Version
Publisher's PDF, also known as Version of record

Specification and estimation of structural econometric models of the labour market: 5 essays

Hans G. Bloemen

Specification and Estimation of Structural Econometric Models of the Labour Market: 5 Essays

Proefschrift

ter verkrijging van de graad van doctor aan de Katholieke Universiteit Brabant, op gezag van de rector magnificus, prof. dr. L.F.W. de Klerk, in het openbaar te verdedigen ten overstaan van een door het college van dekanen aangewezen commissie in de aula van de Universiteit op vrijdag 21 januari 1994 te 14.15 uur door

Hans Gerald Bloemen

geboren te Hengelo.

Promotor: prof. dr. ir. A. Kapteyn

Preface

Chapter 2 of this thesis is virtually identical to Bloemen and Kapteyn (1993a). The same holds for chapter 3 and Bloemen and Kapteyn (1993b). Data were provided by the Netherlands Organization for Strategic Labor Market Research (OSA) and the Netherlands Central Bureau of Statistics (CBS).

Acknowledgments

This thesis has benefitted from the support of many people. I am much indebted to Arie Kapteyn for giving me the opportunity to write this thesis, for his guidance, and for believing in me so early. The contents of this thesis has benefitted from the remarks of various persons. For this I would like to thank Arthur van Soest, Bertrand Melenberg, Tony Lancaster, Vassilis Hajivassiliou and Marcel Kerkhofs. Finally, I would like to thank everbody at CentER, the department of econometrics and the faculty and those outside the university, who, in one way or another, gave me support in the last few years: I will not mention everybody personally: You know who you are!

Contents

Chapter 1

Introduction

The five essays in this dissertation deal with microeconometric models of individual labour market behaviour. Individual decision makers try to behave optimally, given the constraints they face. The stochastic specification of the models is such that a close link with the behavioural relation derived from the economic models is maintained. Different models of the labour market are studied, ranging from static models of labour supply, which explain the labour supply decision at a given point in time, to models in which transitions between different labour force states are incorporated. The static models can be subdivided into models that are neo-classical and models in which the individual is faced by demand side restrictions. In four of the five essays, we use estimation by simulation methods. Estimation by simulation methods is useful if either multidimensional probability integrals appear in the likelihood function, or the region of integration is defined only implicitly, or integrals with integrands that are costly to evaluate appear. The five studies, chapters 2 through 6 in this thesis, all start with an introduction of their own. Therefore, this introductory chapter is kept brief and emphasis is placed on the common aspects and interrelations between the separate papers.

In chapter 2, a method of simulated scores (MSS) estimator for models with limited dependent variables is considered. Three variants of the MSS method are applied to a simple neo-classical labour supply model that consists of a wage equation and a labour supply function. The simplicity of the model enables the estimation of the model by maximum likelihood (ML) and the comparison of the ML estimator with the MSS estimators. To investigate the performance of the estimators, the methods are applied to Monte Carlo data. The MSS estimators turn out to perform in a satisfactory manner, even with a limited number of drawings. When applying the methods of estimation to real data, considerable differences between the parameter estimates obtained by different models, as well as the estimates of wage and participation elasticities are revealed. This is attributed to the oversimplification of the economic model.

In chapter 3 one of the MSS methods described in chapter 2 is applied to a more sophisticated neo-classical model of labour supply. This model is a utility consistent static labour supply model with flexible preferences, a non-linear and possibly non-convex budget set and a wage equation. Three random error terms are included to represent respectively optimization and reporting errors, stochastic preferences, and heterogeneity in wages. Coherency conditions on parameters and supports of error distributions are imposed for all observations. The complexity of the model makes it impossible to write down the integration region of the probability of participation explicitly. As a solution to this problem, frequency simulators can be used to simulate participation probabilities. The most generally applicable frequency simulator is the simple frequency simulator, but by exploiting specific properties of the model at hand, simulators that are combinations of simple frequency simulators and smooth simulators may be employed. The properties of the estimation method adopted are first investigated by using Monte Carlo data. After that the model is estimated for Dutch data. The approach is compared with various simpler alternatives proposed in the literature. It turns out that both in the Monte Carlo data and in the real data the various estimation methods yield very different results. Moreover, since the MSS estimation method yields good results for the Monte Carlo data, it is suggested that the simplifications adopted in the literature may have generated considerable biases.

The labour supply models specified in chapters 2 and 3 are neo-classical in the sense that no allowance is made for structural differences between observed labour supply and optimal labour supply. In reality, the decision maker is often faced by job constraints and hours contraints. In the literature, a series of articles has appeared which deal with the behaviour of individuals who are faced by hours constraints in a static context. (Dickens and Lundberg (1985), Tummers and Woittiez (1991), Van Soest, Woittiez and Kapteyn (1990)). In these models, individuals receive a random number of job offers (possibly zero) at a given point in time, which all have the same gross wage rate, but which may differ in the number of weekly working hours. The job offer which yields the highest level of utility is chosen and its utility level is compared with the utility of non-participation, after which the participation decision is made. To provide a link with job search theory, in chapter 4 this type of model is extended by allowing the wage rate to vary across job offers as well. Moreover, the probability of receiving a job offer is made dependent on individual characteristics. The estimation results reveal that it may be desirable to use, apart from data on labour supply and wages, information that is related to the number of job offers that has been received. The recommended alternative approach is to extend the model by assuming that job offers arrive sequentially over time instead of arriving at a given point in time. In the estimation of this extended model, the availability of data on unemployment duration is required.

In chapter 5 elaborates this approach on. The standard job search model, in which a job is characterized by a wage rate only, is extended by assuming that job offers are characterized by both wage and number of working hours, which arrive jointly from a job offer distribution. The model allows for unobserved heterogeneity in preferences and arrival rates. Two specifications are considered which differ with respect to the way labour supply enters the model. The first model is neo-classical in the sense that once a job offer has arrived, the individual can determine optimal hours by himself. In the second model job searchers are offered fixed packages of wages and working hours. Residual analysis reveals that the latter model outperforms the first. Simulation methods are used to integrate out unobserved heterogeneity.

In chapter 6 an empirical job search model is presented in which the individual decision maker can determine his job offer arrival rate by varying the intensity of search. The availability of several indicators of search intensity allows us to identify cost of search parameters and to examine the effectiveness of search in terms of labour force state duration. The model describes the behaviour of unemployed and employed individuals. Three types of labour market transitions may occur: transitions from unemployment into employment, job to job transitions and transitions from employment into unemployment. Apart from the search intensity indicators, data on unemployment duration and job tenure are used in the estimation of a structural model of job search.

A brief summary and evaluation of the various results is given in chapter 7.

Chapter 2

The joint estimation of a non-linear labour supply function and a wage equation using simulated response probabilities

2.1 Introduction

This chapter investigates the applicability of estimation methods for labour supply models which make use of simulators for the response probabilities. The methods of estimation usually applied, like maximum likelihood and the method of moments, make use of the probabilities of individuals participating in the labour force. If the labour supply model is non-linear and if one wants to incorporate the tax and social security system, thereby assuming that the budget constraints of the individuals are non-convex, the calculation of the participation probabilities may be impossible and there is probably no other way to estimate the model other than making use of simulated moments types of estimators.

McFadden (1989) presents a method of simulated moments estimator for the multinomial response model. The attractiveness of the method is that the number of replications which is used to simulate the response probabilities can be kept fixed to any positive integer without destroying the consistency property of the estimator. In a short time an extensive literature has blossomed in which this approach has been extended and refined. The emphasis has been on computational accuracy and speed in the evaluation of multi-dimensional probabilities, often under normality or closely related assumptions and linearity. See Hajivassiliou (1992) for an overview.

The method of simulated moments estimator, however, is in its simplest form only suitable for discrete response data, whereas in labour supply models the data are usually of a mixed discrete-continuous nature, giving information on whether or not individuals are working and if so, how many hours. Furthermore, any realistic utility consistent model will entail non-linearities and non-normality, so that the various refinements mentioned will not be applicable. Therefore, we want to set up estimation by simulation methods for the mixed discrete-continuous type of model that one typically finds in labour supply analysis.

Different routes can be followed. The most straightforward way is to replace the response probabilities in the likelihood function by simulators, thereby simulating the likelihood function. It can be inferred from Gouriéroux and Montfort (1989) that this method of simulated maximum likelihood (SML) is not consistent for an arbitrary, fixed number of replications to simulate the response probabilities. Lerman and Manski (1981) show that a similar method for the multinomial response model may require huge numbers of replications. To circumvent this problem an alternative is to use the method of simulated scores (MSS) which is based on the simulation of the vector of scores of the log-likelihood function. As in McFadden (1989), the point of departure is the property of the likelihood function that under weak regularity conditions the expectation of the score vector equals zero at the true parameter value. This score vector will be replaced by a simulated score vector. An estimator can be obtained by minimizing the length of the simulated score. The score vector will be simulated in such a way that the property of having a zero expectation at the true parameter value carries over to the simulated score vector. There is no unique way to achieve this and therefore we will propose and compare three different methods of estimation. The method of simulated scores is also used by Hajivassiliou (1989).

The methods of estimation will be applied to the joint estimation of a labour supply function, non-linear in the wage rate, and a wage equation. In this application, we assume a linear budget constraint. This is a rather simple model and in fact it can be estimated by maximum likelihood, using numerical integration. The main purpose of the application is to gain insight into the practical properties of the MSS methods. Hence, we have chosen a model simple enough so that ML is feasible and we can compare the performance of the MSS estimators with ML. In chapter 3 one of the MSS estimators is applied to a much more complicated model with random preferences and non-convex budget constraints. In that model ML is not feasible.

We will present Monte Carlo results as well as real data estimates. In the Monte Carlo study different MSS estimators are compared with each other as well as with ML and SML with a limited number of replications.

The order of presentation is as follows: In the next section we set out the basic model where labour supply is a (possibly non-linear) function of the wage rate and non-labour income. Errors in the wage equation are additive and their nature remains unspecified. In section 2.3 different simulators for the score vector are proposed, each of them generating an alternative method of estimation. Attention is paid to the statistical properties of the methods. In Section 2.4 the properties of these estimators and of ML are first investigated by means of some Monte Carlo experiments. Next, the estimators are applied to the analysis of labour supply of Dutch females.

The general finding is that the MSS estimators perform quite well, though of course slightly below ML, whereas simulated ML may perform poorly. We conclude that the MSS estimators proposed present viable routes for the estimation of utility consistent labour supply models.

2.2 The basic model

Our point of departure is a two-equation model consisting of a labour supply equation and a log-wage equation.

$$h_n^* = h(w_n, \mu_n; \beta) + \epsilon_n \tag{2.1}$$
$$\log(w_n) = w(x_n, \eta) + u_n \tag{2.2}$$

h_n^* is observed and equals the number of hours worked by individual n if the n-th individual is working; h_n^* is unobserved if the n-th individual is non-working,

$$h_n = h_n^* \text{ if } h_n^* > 0 \tag{2.3}$$
$$h_n = 0 \text{ if } h_n^* \leq 0 \tag{2.4}$$

where h_n is the actual number of hours worked by individual n, w_n is the after-tax real wage rate which will be unobserved for a non-working individual, μ_n is non-labour income, x_n is a vector of observable characteristics of individual n, β is a parameter vector with dimension l, η is a vector of parameters with dimension q, ϵ_n and u_n are random disturbances with expectation 0 and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon u} \\ \sigma_{\epsilon u} & \sigma_u^2 \end{pmatrix}$$

and joint probability density function $f(\epsilon_n, u_n)$, independent across observations. For ease of notation we introduce the dummy variable d_n with

$$d_n = 1 \text{ if } h_n^* \leq 0 \tag{2.5}$$
$$d_n = 0 \text{ if } h_n^* > 0 \tag{2.6}$$

We start by deriving the joint probability density function of h_n and w_n given x_n and μ_n. First, an expression for the joint density of h_n^* and w_n has to be found. Using the 1-1 transformations $\epsilon_n = h_n^* - h(w_n, \mu_n; \beta)$ and $u_n = \log(w_n) - w(x_n; \eta)$, we can employ the joint density of ϵ_n and u_n to get the density function $g^*(h_n^*, w_n)$ of h_n^* and w_n. The Jacobian of the transformation is

$$\begin{vmatrix} \frac{\partial \epsilon_n}{\partial h_n^*} & \frac{\partial u_n}{\partial h_n^*} \\ \frac{\partial \epsilon_n}{\partial w_n} & \frac{\partial u_n}{\partial w_n} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -\frac{\partial h}{\partial w_n} & \frac{1}{w_n} \end{vmatrix} = \frac{1}{w_n}$$

So

$$g^*(h_n^*, w_n) = f(h_n^* - h(w_n, \mu_n; \beta), log(w_n) - w(x_n; \eta))\frac{1}{w_n} \tag{2.7}$$

$$-\infty < h_n^* < \infty$$
$$0 < w_n < \infty$$

From this we can derive the mixed discrete-continuous probability density function of h_n and w_n, $g(h_n, w_n|x_n, \mu_n, \theta)$ where θ contains the parameters of β, η and the upper triangular or, equivalently, the lower triangular elements of Σ.

$$g(h_n, w_n|x_n, \mu_n, \theta) = \begin{cases} P(h_n^* \leq 0|x_n, \mu_n, \theta) & \text{if } h_n = 0 \\ g^*(h_n, w_n|x_n, \mu_n, \theta) & \text{if } h_n > 0, 0 < w_n < \infty \end{cases}$$

where

$$P(h_n^* \leq 0|x_n, \mu_n, \theta) = \int_0^\infty \int_{-\infty}^0 g^*(h, w|x_n, \mu_n, \theta)dhdw \tag{2.8}$$

For ease of notation this probability will be denoted by $P_n(\theta)$ or by P_n. The wage w is integrated out because for non-working individuals we have no observations on the wage rate. We shall denote the probability of working $1 - P_n(\theta)$ by $\bar{P}_n(\theta)$ or simply by $\bar{P}_n$. We assume that our sample is ordered in such a way that the observations 1 to N_1 refer to non-working individuals and the observations $N_1 + 1$ to N are working individuals.

We now formulate the log-likelihood function of the model.

$$L(\theta|x_n, \mu_n, w_n, h_n, n = 1, ..., N) =$$

$$\sum_{n=1}^{N_1} \ln P_n(\theta) + \sum_{n=N_1+1}^{N} \ln g^*(h_n, w_n|x_n, \mu_n, \theta) \tag{2.9}$$

This is differentiated with respect to θ to derive the first order conditions for a maximum

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{n=1}^{N_1} \frac{\partial \ln P_n(\theta)}{\partial \theta} + \sum_{n=N_1+1}^{N} \frac{\partial \ln g^*(h_n, w_n|x_n, \mu_n, \theta)}{\partial \theta} \tag{2.10}$$

$$\frac{\partial L(\hat{\theta}_{ML})}{\partial \theta} = 0 \tag{2.11}$$

where $\hat{\theta}_{ML}$ is the maximum likelihood estimator of θ.

Alternatively, we can rewrite the derivative of the log-likelihood function as

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{n=1}^{N} \left[d_n \frac{\partial \ln P_n(\theta)}{\partial \theta} + (1 - d_n) \frac{\partial \ln g^*(h_n, w_n|x_n, \mu_n, \theta)}{\partial \theta} \right] \tag{2.12}$$

where d_n is the dummy variable introduced above.
Let θ_0 be the true parameter value. It is easy to show that if the supports of h_n and w_n do not depend on θ,

$$E\left(\frac{\partial L(\theta_0)}{\partial \theta}\right) = 0 \tag{2.13}$$

which is the result of the fact that the expectation of the derivative of the log-density function with respect to θ at the true parameter value equals zero. In fact, the first order derivative of the log-likelihood function divided by the sample size can be looked upon as a moment estimator of

$$E\left(\frac{\partial \ln g(h, w|x, \mu, \theta)}{\partial \theta}\right) \tag{2.14}$$

evaluated at the parameter value θ.

A procedure which is often followed in estimating this type of model is a two-step procedure. First, the wage equation is estimated using data on working individuals. The resulting estimates of the parameter vector η and the characteristics x of the non-working individuals are used to construct a proxy for the wage variables of the non-working individuals. Second, this constructed data-set is used to estimate the labour supply function, using Tobit-like methods. This method will in general yield inconsistent estimates, particularly for non-linear labour-supply functions. A correct procedure would be to estimate the model simultaneously. A drawback of simultaneous estimation of the model is the difficulty in calculating the response probabilities analytically whenever the model is non-linear, whereas numerical approximation can be expensive and time-consuming. Our purpose is to develop estimation methods for models of the type (1.1)-(1.2) that allow for simultaneous estimation also if the function h is quite complicated (as for instance in the case, where h represents the outcome of utility maximization under a non-linear and non-convex budget constraint). To this end we make use of simulators for the response probabilities like the ones proposed by McFadden (1989). However, unlike the method used in his paper, we shall not restrict ourselves to the discrete response model. We will employ a discrete-continuous type of estimation method. The property (2.13) will be used to develop a method of simulated moments type of estimation. We want to replace the response probabilities in (2.12) by simulators and we want to do that in such a way that property (2.13) carries over to the simulated score vector. Then an estimator can be found by minimizing the length of the simulated score vector.

2.3 Estimation

Three ways of simulating the score of the log-likelihood function are considered. The first method replaces the discrete part of the score by an expression with an instrument matrix and a simulator for the response probability of non-working individuals. The disadvantage of this method is that the consistency of the estimator will depend on the choice of the matrix of instruments. This is due to the fact that the expectation of the simulated score, evaluated at the true parameter value θ_0, does not equal zero, unless a specific form for the matrix of instruments is chosen, which makes use of a consistent estimator for θ_0. So, in order to get a consistent estimator, we need a consistent estimator obtained from a different estimation procedure. Therefore, the first method is only useful to increase efficiency of the first round estimator obtained by one of the next two methods. The second method of estimation also uses a matrix of instruments. The estimator will be consistent, irrespective of the choice of the instruments. To simulate the score, simulators of the response probabilities and their derivatives are needed for each individual, both non-working and working. A second estimation round can be performed to increase the efficiency of the estimators, using an updated version of the matrix of instruments. This method directly extends Mc Fadden's (1989) estimation method for the discrete response model, by adding a continuous component to his objective function. The third method does not rely on a matrix of instruments. Only simulators of the derivatives of the response probabilities are required.

2.3.1 Method 1

We rewrite the first order derivative of the log-likelihood function in the following way.

$$\frac{\partial L(\theta)}{\partial \theta} = \sum_{n=1}^{N} \left[d_n Z_n (1 - P_n) + (1 - d_n) \frac{\partial \ln g^*(h_n, w_n | x_n, \mu_n, \theta)}{\partial \theta} \right] \tag{2.15}$$

$$\text{where} \quad Z_n := \frac{\frac{\partial P_n}{\partial \theta}}{P_n(1-P_n)} \tag{2.16}$$

Now the vector Z_n is replaced by an arbitrary vector of instruments $\bar{Z}_n$, which does not depend on θ. The resulting expression is

$$\frac{\partial \bar{L}^1(\theta)}{\partial \theta} = \sum_{n=1}^{N} \left[d_n \bar{Z}_n (1-P_n) + (1-d_n) \frac{\partial \ln g^*(h_n, w_n | x_n, \mu_n, \theta)}{\partial \theta} \right] \tag{2.17}$$

where the superscript 1 refers to the number of the method. Calculating the expectation of expression (2.17), conditional on $\bar{Z}_n$, evaluated at the true parameter value θ_0 yields

$$E\left(\frac{\partial \bar{L}^1}{\partial \theta} | \bar{Z}_n \right) = \sum_{n=1}^{N} \left[P_n \bar{Z}_n (1-P_n) + \frac{\partial \bar{P}_n}{\partial \theta} \right] \tag{2.18}$$

(with $\bar{P}_n = 1 - P_n$) which in general doesn't equal zero. However, if we construct $\bar{Z}_n$ in such a way that

$$plim(\bar{Z}_n) = \frac{\frac{\partial P_n}{\partial \theta}}{P_n(1-P_n)} \text{ at } \theta_0 \tag{2.19}$$

we see that at the true parameter value θ_0 the simulated score has asymptotic mean equal to zero.

The response probability $P_n(\theta)$ in expression (2.17) is replaced by a frequency simulator or by a so called smooth unbiased simulator $k_n(\theta, v_R^*)$ where v_R^* is a vector of R drawings from some distribution which does not depend on θ. The simulator is unbiased if it has the property

$$E(k_n(\theta, v_R^*)) = P_n(\theta) \tag{2.20}$$

A smooth unbiased simulator can be constructed in the same way as in Monte Carlo importance sampling, see e.g. Hammersley and Handscomb (1964). Take a density function $\gamma(h, w)$ with support coinciding with the bounds of the integrals in expression (2.8) for P_n. The response probability can be rewritten as:

$$P_n(\theta) = \int_0^{\infty} \int_{-\infty}^{0} \tau(h, w | x_n, \mu_n, \theta) \gamma(h, w) dh dw \tag{2.21}$$

with

$$\tau(h, w | x_n, \mu_n, \theta) = \frac{g^*(h, w | x_n, \mu_n, \theta)}{\gamma(h, w)} \tag{2.22}$$

For every n, R random vectors $(h_{(r,n)}, w_{(r,n)})$ are drawn from the density function $\gamma(h, w)$, independently across observations and not depending on the parameter vector θ. By averaging over the drawings a simulator is obtained:

$$k_n(\theta, v_R^*) = \frac{1}{R} \sum_{r=1}^{R} \tau(h_{(r,n)}, w_{(r,n)} | x_n, \mu_n, \theta) \tag{2.23}$$

where v_R^* consists of the drawings $(h_{(r,n)}, w_{(r,n)})$. The function $\tau(.,.|.,.,.)$ is the so called weight function which corrects for the fact that we are drawing random numbers from the distribution with density function $\gamma(.,.)$ instead of the true distribution with density function $g^*(.,.|.,.,.)$. If $\gamma(.,.)$ and $g^*(.,.|.,.,.)$ coincide the weight function is identically equal to 1. In our application, described in the next section, we will actually use a slightly different way of simulating the response probabilities by exploiting the normality assumptions and the assumed linearity of the log-wage equation. Therefore, we don't have to choose a density function $\gamma(h,w)$. However, in more complicated applications, as in chapter 3, weight functions are necessary.

As indicated by McFadden (1989), a simulator for the derivatives of P_n can be constructed in a similar way. An unbiased simulator $m(\theta, v_R^*)$ of the derivatives of P_n is

$$m_n(\theta, v_R^*) = \frac{1}{R}\sum_{r=1}^{R} \frac{\partial \tau(h_{(r,n)}, w_{(r,n)}|x_n, \mu_n, \theta)}{\partial \theta} \tag{2.24}$$

For the simulation of (2.17) we only need k_n. The simulated score is:

$$K_R^1(\theta) = \sum_{n=1}^{N} \left[d_n \bar{Z}_n(1 - k_n(\theta, v_R^*)) + (1-d_n)\frac{\partial \ln g^*(h_n, w_n|x_n, \mu_n, \theta)}{\partial \theta} \right] \tag{2.25}$$

The estimation procedure now becomes: Minimize the length of the simulated score vector:

$$\min_{\theta} s_N(\theta) \quad \text{where} \tag{2.26}$$
$$s_N(\theta) \quad = \quad K_R^1(\theta)' K_R^1(\theta) \tag{2.27}$$

where the matrix of instruments has to be based on the following formula, evaluated in the true parameter point (or at least at a consistent estimate of it):

$$\bar{Z}_n = \frac{m_n(\theta, v_{R_Z}^{**})}{k_n(\theta, v_{R_Z}^{**})(1 - k_n(\theta, v_{R_Z}^{**}))} \tag{2.28}$$

where $m_n(\theta, v_{R_Z})$ is an unbiased simulator of the derivatives of $P_n(\theta)$, and $v_{R_Z}^{**}$ are drawn independently of the v_R^* which are used in minimizing $s_N(\theta)$. The number of drawings to simulate the vector of instruments is denoted by R_Z to indicate that it is not necessarily equal to R, the number of drawings used to construct $k_n(\theta, v_R^*)$ in (2.25). The reason for presenting this method of estimation is that the asymptotic variance of this method is lower than the aymptotic variance of the methods presented in the next two subsections, which will be explained in section 2.3.4. Therefore, a two-step procedure could be followed: First, obtain a consistent estimate by applying one of the other estimation methods and second, use the estimates to construct the vector of instruments in (2.28) and apply method 1 to increase the efficiency of the estimates.

We are interested in the error we make by replacing the score vector by a simulator. Therefore, the simulated score in (2.25) is rewritten as the sum of the true score in (2.15) and a simulation residual.

$$K_R^1(\theta) = \frac{\partial L}{\partial \theta} + RES_1 \tag{2.29}$$

where

$$RES_1 = \sum_{n=1}^{N} d_n Z_n (P_n - k_n) + \sum_{n=1}^{N} d_n (\bar{Z}_n - Z_n)(1 - k_n) \tag{2.30}$$

For ease of notation the arguments are omitted. The first term of (2.30) can be rewritten as

$$\frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R} d_n Z_n (P_n - k_{nr}) \tag{2.31}$$

where k_{nr} is the r-th term of k_n. By increasing the number of drawings R this term will tend to zero. In the second term, the vector $\bar{Z}_n$ appears which is a non-linear function of the simulators. Because the two factors in this term are independent by construction, we concentrate on $\bar{Z}_n - Z_n$. If $\bar{Z}_n$ is constructed in a proper way, i.e. by using a consistent estimate for θ, then

$$\text{plim}_{R_Z \to \infty}(\bar{Z}_n - Z_n) = 0 \tag{2.32}$$

so, by taking the number of drawings to construct the vector of instruments large enough the second term can also be made arbitrarily small. The variance of the simulation residual determines the loss of efficiency caused by using the simulated score instead of the true score vector. In the next subsection we discuss how the variance of the simulation residual can be influenced by the number of drawings R.

Since the derivation of the asymptotic distribution is equivalent for the three estimators we will treat the asymptotic properties of the estimators at the end of this section.

2.3.2 Method 2

In order to obtain the second method of estimation we rewrite the score of the log-likelihood function as

$$\frac{\partial L}{\partial \theta} = \sum_{n=1}^{N}\left\{ Z_n(d_n - P_n) + (1-d_n)\left[\frac{\partial \ln g^*(h_n, w_n | x_n, \mu_n, \theta)}{\partial \theta} - \frac{\partial \ln \bar{P}_n}{\partial \theta}\right]\right\} \tag{2.33}$$

with Z_n defined as above. The first component of this expression equals the score of the log-likelihood of the binary response model. If we replace the vector Z_n by an arbitary vector of instruments $\bar{Z}_n$, independent of θ, the expectation of the resulting expression, conditional on $\bar{Z}_n$, equals zero at the true parameter value θ_0.

To simulate $P_n(\theta)$ we use the simulator $k_n(\theta, v_R^*)$ defined above. The problem is how to simulate $\frac{\partial \ln \bar{P}_n}{\partial \theta}$. To see why this is a problem we rewrite this expression as

$$\frac{\partial \ln \bar{P}_n}{\partial \theta} = \frac{1}{\bar{P}_n}\frac{\partial \bar{P}_n}{\partial \theta} \tag{2.34}$$

It is not difficult to construct unbiased simulators of $\bar{P}_n$ and $\frac{\partial \bar{P}_n}{\partial \theta}$, as we have shown before. However, if we use their simulators $\bar{k}_n(\theta, v_R^*)$ and $\bar{m}_n(\theta, v_R^*)$ to simulate $\frac{\partial \ln \bar{P}_n}{\partial \theta}$, we don't get un unbiased simulator.

$$E\left(\frac{\bar{m}_n(\theta, v_R^*)}{\bar{k}_n(\theta, v_R^*)}\right) \neq \frac{\partial \ln \bar{P}_n}{\partial \theta} \tag{2.35}$$

So, the expectation of the simulated score evaluated at θ_0 won't equal zero. It is not clear how to get an unbiased simulator of $\frac{\partial \ln \bar{P}_n}{\partial \theta}$. In order to solve this problem, we will simulate $(1-d_n)\frac{\partial \ln \bar{P}_n}{\partial \theta}$ instead of $\frac{\partial \ln \bar{P}_n}{\partial \theta}$.

$$E\left[(1-d_n)\frac{\partial \ln \bar{P}_n}{\partial \theta}\right] = \frac{\partial \bar{P}_n}{\partial \theta} \tag{2.36}$$

We now replace $(1-d_n)\frac{\partial \ln \bar{P}_n}{\partial \theta}$ by $\frac{\partial \bar{P}_n}{\partial \theta}$. As a result, the original score vector is replaced by

$$\frac{\partial \bar{L}^2}{\partial \theta} = \sum_{n=1}^{N}\left[\bar{Z}_n(d_n - P_n) + (1-d_n)\frac{\partial \ln g^*(h_n, w_n|x_n, \mu_n, \theta)}{\partial \theta} - \frac{\partial \bar{P}_n}{\partial \theta}\right] \tag{2.37}$$

Inserting the simulators for the response probabilities and their derivatives in this expression gives the simulated score:

$$K_R^2(\theta) = \sum_{n=1}^{N}\left[\bar{Z}_n(d_n - k_n(\theta, v_R^*)) + (1-d_n)\frac{\partial \ln g^*(h_n, w_n|x_n, \mu_n, \theta)}{\partial \theta} - \bar{m}_n(\theta, v_R^*)\right] \tag{2.38}$$

The estimation procedure becomes: Choose instrument vectors $\bar{Z}_n$ and minimize the length of the simulated score.

$$\min_{\theta} K_R^2(\theta)' K_R^2(\theta) \tag{2.39}$$

We can increase the efficiency of the estimator by rerunning the procedure using the updated vectors of instruments, as described in the preceding subsection.

Again we are interested in the simulation residual. First, (2.33) with Z_n replaced by $\bar{Z}_n$ is compared with (2.38). Then the following residual is obtained:

$$\frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R}\left[\bar{Z}_n(P_n - k_{nr}) - \left\{\bar{m}_{nr} - (1-d_n)\frac{\partial \ln \bar{P}_n}{\partial \theta}\right\}\right] \tag{2.40}$$

This is the simulation residual corresponding to the comparison of the moments estimator (Z_n replaced by $\bar{Z}_n$) with the simulated moments estimator. The dummy variable can be rewritten as

$$\begin{aligned} d_n &= P_n + \nu_n \\ \text{with } E(\nu_n) &= 0 \\ \text{and } Var(\nu_n) &= P_n(1-P_n) \end{aligned} \tag{2.41}$$

Inserting this in the residual gives:

$$\frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R}\left[\bar{Z}_n(P_n - k_{nr}) - \left\{\bar{m}_{nr} - \frac{\partial \bar{P}_n}{\partial \theta}\right\}\right] - \sum_{n=1}^{N}\nu_n\frac{\partial \ln \bar{P}_n}{\partial \theta} \tag{2.42}$$

The variance of the first term of (2.42) can be reduced by increasing the number of drawings R. Suppose that $Var[\bar{Z}_n(P_n - k_{nr}) - \{\bar{m}_{nr} - \frac{\partial \bar{P}_n}{\partial \theta}\}] = \Xi_n$, conditional on the instruments $\bar{Z}_n$. This variance does not depend on r because the drawings are i.i.d. Then the variance of the first term is $\frac{1}{R}\sum_{n=1}^{N}\Xi_n$. With fixed N, increasing R to infinity results in reducing this variance to zero. The second term is the error which is caused by the fact that $(1-d_n)\frac{\partial \ln \bar{P}_n}{\partial \theta}$ is simulated by a simulator for $\frac{\partial \bar{P}_n}{\partial \theta}$. The expectation of this term equals zero, whereas the variance equals $\sum_{n=1}^{N} \bar{P}_n P_n \frac{\partial \ln \bar{P}_n}{\partial \theta} \frac{\partial \ln \bar{P}_n}{\partial \theta'}$. This term of the simulation residual cannot be influenced by the number of drawings. Therefore, this term leads to inefficiency, also for large R.

To compare the efficiency of the method of simulated moments estimator to the maximum likelihood estimator, also the term involving the difference between $\bar{Z}_n$ and Z_n has to be taken into account. The simulation residual then becomes:

$$RES_2 = \frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R}\left[Z_n(P_n - k_{nr}) - \left\{\bar{m}_{nr} - \frac{\partial \bar{P}_n}{\partial \theta}\right\}\right] + \sum_{n=1}^{N}(\bar{Z}_n - Z_n)(d_n - k_n) - \sum_{n=1}^{N}\nu_n \frac{\partial \ln \bar{P}_n}{\partial \theta} \tag{2.43}$$

Here the same observations can be made as for method 1.

2.3.3 Method 3

As opposed to the first two methods of estimation, we don't rewrite the score of the log-likelihood function in a form involving instrument vectors Z_n. Instead, we immediately replace the discrete part of the score by a simulator. In analogy with the preceding method, we replace $d_n \frac{\partial \ln P_n}{\partial \theta}$ by $\frac{\partial P_n}{\partial \theta}$. This yields the expression

$$\frac{\partial \bar{L}^3}{\partial \theta} = \sum_{n=1}^{N}\left[\frac{\partial P_n}{\partial \theta} + (1 - d_n)\frac{\partial g^*(h_n, w_n|x_n, \mu_n, \theta)}{\partial \theta}\right] \tag{2.44}$$

The derivative of the probability is replaced by its simulator $m_n(\theta, v_R^*)$, which leads to the following expression for the simulated score:

$$K_R^3(\theta) = \sum_{n=1}^{N}\left[m_n(\theta, v_R^*) + (1 - d_n)\frac{\partial g^*(h_n, w_n|x_n, \mu_n, \theta)}{\partial \theta}\right] \tag{2.45}$$

The estimation procedure becomes

$$\min_{\theta} K_R^3(\theta)' K^3(\theta) \tag{2.46}$$

We compare this method of simulating the score with the method described in the preceding subsection. Because $\bar{P}_n = 1 - P_n$, we find that

$$\frac{\partial \bar{P}_n}{\partial \theta} = -\frac{\partial P_n}{\partial \theta} \tag{2.47}$$

The same relation holds for their simulators:

$$m_n(\theta, v_R^*) = -\bar{m}_n(\theta, v_R^*) \tag{2.48}$$

Inserting this in the simulated score, we obtain

$$K_R^3 = \sum_{n=1}^{N}\left[(1 - d_n)\frac{\partial g^*(h_n, w_n|x_n, \mu_n, \theta)}{\partial \theta} - \bar{m}_n(\theta, v_R^*)\right] \tag{2.49}$$

This is exactly the second component of the simulated score of method 2. Obviously, by replacing the derivative of the log-response probability by a simulator for the derivative of the response probability, some information is lost and $\frac{\partial \bar{L}^3}{\partial \theta}$ and the second component of $\frac{\partial \bar{L}^2}{\partial \theta}$ become indistinguishable. Method 2 has the intuitively appealing property that it includes the minimization of the distance between the binary response indicator d_n and its theoretical expectation P_n. The simulation residual for method 3 can be found in the same way as for method 2 and is given by

$$RES_3 = \frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R}\left[m_{nr} - \frac{\partial P_n}{\partial \theta}\right] - \sum_{n=1}^{N}\nu_n \frac{\partial \ln \bar{P}_n}{\partial \theta} \tag{2.50}$$

The residual contains the same error which is not influenced by the number of drawings R as method 2.

2.3.4 Asymptotic distribution of the estimators

In the preceding subsections we presented three ways to simulate the first order derivatives of the log-likelihood function. The simulated score vectors of methods 2 and 3 satisfy the property that their expectation, evaluated at the true parameter vector θ_0, equals zero, whereas the probability limit of the simulated score of method 1 equals zero at the true parameter value θ_0 if the vectors of instruments are constructed in a proper way. Therefore, the length of the expectation of the simulated score is minimized at the true parameter value θ_0. It is intuitively clear that if we minimize the length of the simulated score, the resulting parameter vector $\hat{\theta}_R^i$ at which minimization takes place, will converge to the true parameter value θ_0, or, equivalently, $\hat{\theta}_R^i$ will be a consistent estimator of θ_0.

We assume that

$$\frac{1}{\sqrt{N}} K_R^i(\theta_0) \;\widetilde{\text{asy}}\; N(0, V_{R0}^i), i = 1, ..., 3 \tag{2.51}$$

with V_{R0}^i some positive definite symmetric matrix. Below, we explain that this assumption can be justified on the basis of central limit arguments. Using this assumption and the consistency of $\hat{\theta}_R^i$, apart from usual regularity assumptions on the parameter space and the like, it is possible to show (see, for example, Pakes and Pollard (1989)) that

$$\sqrt{N}(\hat{\theta}_R^i - \theta_0) \;\widetilde{\text{asy}}\; N(0, (\Gamma^{i'}\Gamma^i)^{-1}\Gamma^{i'}V_{R0}^i\Gamma^i(\Gamma^{i'}\Gamma^i)^{-1}) \tag{2.52}$$

$$\text{where } \Gamma^{i'} = plim \frac{1}{N}\frac{\partial(\frac{\partial \bar{L}^i(\theta_0)}{\partial \theta'})}{\partial \theta} \tag{2.53}$$

We will now comment on the assumed normality of the simulated score. The expectation of the simulated score, conditional on the instruments, equals zero at θ_0 for $i = 2, 3$. It was shown that the simulated score could be written as the sum of the true score vector and a simulation residual. It is well-known that under general conditions the distribution of the true score vector divided by the square root of the sample size converges to a normal distribution. The assumed independence across observations and the fact that the simulators are constructed using independent random drawings can be used to apply the Lindeberg-Feller central limit theorem to prove the normality of the simulation residuals (conditional on the instruments) divided by the square root of the sample size. The distribution of the simulated score then converges to the sum of two normal distributions which is in turn a normal distribution. We can do the same for method 1, but not without recalling that the instruments have to be constructed such that (2.19) is satisfied.

Using the expression of the asymptotic covariance matrix and the results of the analysis of the simulation residuals in the preceding subsections it is possible to analyse the efficiency of the estimators by comparing the asymptotic covariance matrices of the simulation estimators with the asymptotic covariance matrix of the maximum likelihood estimator. It is a well-known result that

$$\sqrt{N}(\hat{\theta}_{ML} - \theta_0) \quad \widetilde{\text{asy}} \quad N(0, \Omega_{ML}) \tag{2.54}$$

$$\text{where } \Omega_{ML} \quad = \quad B^{-1} \tag{2.55}$$

$$B = -plim \frac{1}{N} \frac{\partial^2 L(\theta_0)}{\partial\theta\partial\theta'} \tag{2.56}$$

To clarify the relationship of the covariance matrix of the ML-estimator with the asymptotic covariance matrix of the simulation estimators we rewrite Ω_{ML} as

$$\Omega_{ML} = (\Gamma'_{ML}\Gamma_{ML})^{-1}\Gamma'_{ML}V_{ML}\Gamma_{ML}(\Gamma'_{ML}\Gamma_{ML})^{-1} \tag{2.57}$$

where

$$\Gamma'_{ML} = plim \frac{1}{N} \frac{\partial\left(\frac{\partial L(\theta_0)}{\partial\theta'}\right)}{\partial\theta} = -B \tag{2.58}$$

which is the equivalent of (2.53), and

$$V_{ML} = plim \frac{1}{N} \sum_{n=1}^{N} \frac{\partial L_n(\theta_0)}{\partial\theta} \frac{\partial L_n(\theta_0)}{\partial\theta'} \tag{2.59}$$

Recall that $V_{ML} = B$. For method 1, if the instruments are constructed properly and if the number of drawings to construct the instruments tend to infinity $\Gamma^{1'}$ and Γ'_{ML} are equivalent. Then the efficiency comparison reduces to comparing V_{ML} with V_R^1 for this method. The difference between these matrices is given by the covariance matrix of the simulation residuals. In section 2.3.1 it has been derived that this variance disappears if R tends to infinity. Therefore, we can conclude that if the matrix of instruments is constructed on the basis of (2.28) and if both the number of drawings to construct this matrix and the number of drawings to simulate the response probabilities tend to infinity, the covariance matrix of the method 1 estimator and the covariance matrix of the maximum likelihood estimator are asymptotically equal. Of course, this result has only theoretical meaning because the reason why we construct these simulation estimators is to be able to keep the number of drawings fixed and small.

To examine the efficiency of method 2 it has to be noted first that because of replacement (2.36), the simulated score $K_R^2(\theta)$ does not tend to the true score if N is fixed and R tends to infinity. Therefore, we first need to establish the relation between Γ'_{ML} and $\Gamma^{2'}$. From (2.33) and (2.37) it is readily established that

$$\Gamma^{2'} - \Gamma'_{ML} =$$

$$-plim\tfrac{1}{N}\textstyle\sum_{n=1}^{N}\left[(\bar{Z}_n - Z_n)\frac{\partial P_n}{\partial\theta} + \frac{\partial Z'_n}{\partial\theta}(d_n - P_n) + \nu_n \frac{\partial\left(\frac{\partial \ln \bar{P}_n}{\partial\theta'}\right)}{\partial\theta} + \frac{1}{\bar{P}}\frac{\partial \bar{P}_n}{\partial\theta}\frac{\partial \bar{P}_n}{\partial\theta'}\right] \tag{2.60}$$

from which only the first three terms equal zero if the instruments are constructed such that (2.19) is satisfied, i.e. according to formula (2.28) with drawings tending to infinity. From the analysis of the simulation residuals it becomes clear that if the matrix of instruments is constructed according to (2.28) with drawings tending to infinity, and if the response probabilities and their derivatives are simulated with R tending to infinity as well, the asymptotic variance of the score of the likelihood function, evaluated in a consistent estimator is exceeded by X, where

$$X = plim \left(\frac{1}{N} \sum_{n=1}^{N} P_n \bar{P}_n \frac{\partial \ln \bar{P}_n}{\partial \theta} \frac{\partial \ln \bar{P}_n}{\partial \theta'} \right) \tag{2.61}$$

which was derived in section 2.3.2. The same expression can be derived for method 3.

Finally, to estimate the covariance matrix we calculate

$$\hat{\Omega}_R^i = (\hat{\Gamma}^{i\prime}\hat{\Gamma}^i)^{-1}\hat{\Gamma}^{i\prime}\hat{V}_R^i\hat{\Gamma}^i(\hat{\Gamma}^{i\prime}\hat{\Gamma}^i)^{-1} \tag{2.62}$$

with

$$\hat{\Gamma}^{i\prime} = \frac{1}{N}\frac{\partial(\frac{\partial L^i(\hat{\theta}_R^i)}{\partial \theta'})}{\partial \theta} \tag{2.63}$$

$$\hat{V}_R^i = \frac{1}{N}\sum_{n=1}^{N} K_{nR}^i(\hat{\theta}_R^i)K_{nR}^i(\hat{\theta}_R^i)' \tag{2.64}$$

where the index n indicates the n-th component of the simulated score. Expression (2.64) can be calculated by simulation.

2.4 Monte Carlo and empirical application

We will now illustrate the properties of the various estimators by making specific assumptions about the form of the labour supply function and the log-wage equation. We then compare the estimation methods by using Monte Carlo methods and then by estimating the specification for a sample of 849 married female individuals, drawn in 1985. We'll assume that the preferences of the individuals are described by a utility function introduced by Hausman and Ruud (1984), which implies a labour supply function quadratic in the wage rate and linear in non-labour income. The wage equation is assumed to be log-linear. The disturbances of the labour supply function and the log-wage equation are assumed to be normally distributed.

The specific form of the labour supply function under the assumption of a linear budget constraint becomes:

$$h(w_n, \mu_n, \beta) = \beta_3 + \mu_n^* \beta_2 + w_n \beta_4 \tag{2.65}$$

$$\text{with } \mu_n^* = \beta_1 + \mu_n + w_n \beta_3 + \frac{1}{2} w_n^2 \beta_4 \tag{2.66}$$

Inserting the expression for μ_n^* in the labour supply function and reparameterizing gives:

$$h(w_n, \mu_n, \alpha) = \alpha_1 + \mu_n \alpha_2 + w_n \alpha_3 + \frac{1}{2} w_n^2 \alpha_4 \text{ with} \tag{2.67}$$

$$\begin{aligned}\alpha_1 &= \beta_3 + \beta_1\beta_2 \\ \alpha_2 &= \beta_2 \\ \alpha_3 &= \beta_2\beta_3 + \beta_4 \\ \alpha_4 &= \beta_2\beta_4\end{aligned}$$

The log-wage equation becomes:

$$\log w_n = \sum_{j=1}^{11} \eta_j x_{nj} + u_n \tag{2.68}$$

where

- $x_{n1} = 1$ for all n,
- $x_{n2} =$
  log of number of persons in individual n's family,
- $x_{n3} =$
  the number of children with age below 6 of individual n,
- $x_{n4} =$ log-age of individual n,
- x_{n5} to x_{n8}
  are dummy indicators for the level of education of individual n,
  where x_{n5} is the lowest level of education (educ1),
  x_{n6} is the second lowest education level (educ2) etc.
  For the highest education level, no dummy indicator is included
- x_{n9} and x_{n10}
  are indicators for the type of education received,
  x_{n9} is a dummy indicator for non-technical and non-commercial types of education (sec1),
  x_{n10} is a dummy indicator for semi-technical and semi-commercial types of education (sec2).
  For technical and commercial types of education no dummy indicator is included,
- $x_{n11} = x_{n4}^2$
- here w_n is the after tax hourly wage rate; labour supply will be measured in hours per week

Thus, the complete model reads

$$h_n^* = \alpha_1 + \mu_n\alpha_2 + w_n\alpha_3 + \frac{1}{2}w_n^2\alpha_4 + \epsilon_n \tag{2.69}$$

$$\log w_n = \sum_{j=1}^{11} \eta_j x_{nj} + u_n \tag{2.70}$$

ϵ_n and u_n are jointly normally distributed with

$$\begin{pmatrix} \epsilon_n \\ u_n \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right) \tag{2.71}$$

$$\Sigma = \begin{pmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon u} \\ \sigma_{\epsilon u} & \sigma_u^2 \end{pmatrix} \tag{2.72}$$

$$\begin{aligned} h_n &= h_n^* \text{ if } h_n^* > 0 & (2.73) \\ h_n &= 0 \text{ if } h_n^* \leq 0 & (2.74) \end{aligned}$$

We now derive an expression for the response probabilities and their derivatives. Under the normality assumptions the joint probability density function of h_n and w_n is

$$g(h_n, w_n | x_n, \mu_n, \theta) =$$

$$P_n(\theta) \text{ if } h_n = 0 \tag{2.75}$$

$$\frac{1}{2\pi w_n} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\begin{pmatrix} h_n - h(w_n, \mu_n, \alpha) \\ \log w_n - \eta' x_n \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} h_n - h(w_n, \mu_n, \alpha) \\ \log w_n - \eta' x_n \end{pmatrix}\right\} \tag{2.76}$$

$$\text{if } h_n > 0, 0 < w_n < \infty$$

where

$$P_n(\theta) = \int_{-\infty}^{\infty} \Phi\left(-\frac{h[\exp(\eta' x_n + \sigma_u v), \mu_n, \alpha] + \rho\sigma_\epsilon v}{\sigma_{\epsilon|u}}\right) \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}v^2\} dv \tag{2.77}$$

with

$$\rho = \frac{\sigma_{\epsilon u}}{\sigma_\epsilon \sigma_u}, \sigma_{\epsilon|u}^2 = \sigma_\epsilon^2(1-\rho^2)$$

and where $\Phi(.)$ is the standard normal distribution function.

We have written the double integral in a form with the well-known function $\Phi(.)$ and an integral over the standard normal variable v. As a result we only need to simulate the second integral over v. Because the standard normal distribution function doesn't depend on the parameters, we don't have to choose a weighting density function as described for the general case in the preceding section. To simulate the probability P_n we can simply draw from the standard normal distribution and compute the expression under the integral sign.

We obtain the following expression for the smooth unbiased simulator of P_n.

$$k_n(\theta, v_R^*) = \frac{1}{R}\sum_{r=1}^{R} \Phi\left(-\frac{h[\exp(\eta' x_n + \sigma_u v_{nr}^*), \mu_n, \alpha] + \rho\sigma_\epsilon v_{nr}^*}{\sigma_{\epsilon|u}}\right) \tag{2.78}$$

where the v_{nr}^* are independent random drawings from the standard normal distribution.

To simulate the derivative of P_n with respect to, say, θ_j we first write

$$\frac{\partial P_n}{\partial \theta_j} = \int_{-\infty}^{\infty} \frac{\partial \Phi\left(-\frac{h[\exp(\eta' x_n + \sigma_u v), \mu_n, \alpha] + \rho\sigma_\epsilon v}{\sigma_{\epsilon|u}}\right)}{\partial \theta_j} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}v^2\} dv \tag{2.79}$$

where the derivative in the integrand can be calculated analytically. Then we simulate it by

$$m_{nj}(\theta, v_R^*) = \frac{1}{R}\sum_{r=1}^{R} \frac{\partial\Phi\left(-\frac{h[\exp(\eta' x_n+\sigma_u v_{nr}^*),\mu_n,\alpha]+\rho\sigma_\epsilon v_{nr}^*}{\sigma_\epsilon|u}\right)}{\partial\theta_j} \tag{2.80}$$

2.4.1 Monte Carlo results

To get an idea about the performance of the estimation methods we have performed some Monte Carlo experiments. First of all, values for the parameters in (2.69), (2.70), (2.71) and (2.72) were chosen. Next, disturbances ϵ_n and u_n, n = 1,...,N, were generated, under the assumptions (2.71), (2.72) with $\sigma_{\epsilon u} = 0$. The characteristics x_{nj} from the sample were used to generate wages w_n, n = 1,...,N. Included are the constant term with parameter η_1, log(age) with parameter η_2 and the square of log(age) with parameter η_3. These wages were used to generate the h_n^*, n = 1,...,N. Making use of (2.73) and (2.74) and the non-labour income series from the sample, we generated the "observations" h_n.

This generated data set was used to estimate the model with different methods of estimation. First of all the model is estimated with maximum likelihood (ML) using numerical integration. The Gauss-Hermite quadrature formula is used with a number of abscissae equal to 16, see e.g. Stroud and Secrest (1966). The second method of estimation is simulated maximum likelihood (SML) with a number of drawings R equal to 10. Here the probabilities are simulated according to (2.78) and inserted in the likelihood function directly. Although this method is inconsistent for a fixed and small number of drawings, it would still be useful if the asymptotic bias were small. Finally, the model is estimated using the three MSS estimators, abbreviated as MSS1, MSS2 and MSS3 below. Two different numbers of drawings to simulate the response probabilities are used, i.e. $R = 1$ and $R = 10$. The matrix of instruments is constructed on the basis of formula (2.27) evaluated in the true parameter point, where the number of drawings to simulate the instruments is $R_Z = 10$ for both MSS1 and MSS2. For MSS1 the model is also estimated with $R_Z = 500$ and $R = 1$.

Table 2.1 presents the results of the Monte Carlo study. The Monte Carlo procedure was repeated 20 times, so that the numbers in the tables refer to means over 20 replications. This modest number of replications has been chosen in view of the rather heavy computational burden of non-linear optimization problems in general. For each parameter the first line presents the mean of the estimates as given by:

$$\bar{\theta}_j = \frac{1}{20}\sum_{i=1}^{20} \hat{\theta}_j^i, \quad j = 1, ..., N \tag{2.81}$$

N is the number of parameters, whereas the subscript j stands for the j-th component of the parameter vector. The second line gives the sample standard deviation of the estimates:

$$SD_j = \sqrt{\frac{1}{20-1}\sum_{i=1}^{20}(\hat{\theta}_j^i - \bar{\theta}_j)^2}, \quad j = 1, ..., N \tag{2.82}$$

The third line presents the relative error:

$$\text{rel. err. } = 100\% \times |\bar{\theta}_j - \theta_{0j}|/|\theta_{0j}| \tag{2.83}$$

From the table it can be seen that the ML estimator in general performs best in the sense that it has the lowest standard errors and the lowest relative errors. It is clear that SML with $R = 10$ performs very badly. Its performance is even worse than all of the three MSS variants with $R = 1$. Therefore, the use of MSS instead of SML is not just something that only has theoretical relevance. By comparing different numbers of drawings of the same MSS estimator it can be seen that the standard deviations decrease with the increase in number of drawings, which is consistent with the analysis of the simulation residuals in section 2.3.1. For MSS1 the number of drawings R_Z to simulate the matrix of instruments has been increased from 10 to 500. Recall that for this method the construction of the matrix of instruments not only affects the efficiency of the estimator but also its consistency. MSS1 with $R = 1$ and $R_Z = 500$ outperforms MSS1 with $R = 10$ and $R_Z = 10$. The standard deviations are comparable with those of MSS2 with $R = 1$ and $R_Z = 10$, but they are not lower than the MSS2 standard deviations, which questions the use of MSS1. Comparing MSS2 with MSS3 it can be said that for $R = 1$ method 2 has the lower standard deviations and the higher relative errors, whereas for $R = 10$ the differences are rather small, although MSS3 seems to perform slightly better, which may be due to the fact that in MSS2 theoretically an additional inefficiency is introduced by simulating the matrix of instruments with only $R_Z = 10$ drawings.

One may question the practical relevance of the experiments so far with respect to MSS1 and MSS2, since the instruments were computed at the true parameter point which of course is unknown in practice. To see how this affects results, Table 2.2 presents results for MSS1 and MSS2 with instruments based on estimates obtained with MSS3. Moreover, Table 2.2 also presents the mean standard error of the estimates over the twenty replications so that a comparison with the sample standard deviations is possible. Comparing the table with the corresponding columns in Table 2.1, one observes that MSS1 is a bit more sensitive to the choice of instruments than MSS2, as one would expect. Actually, MSS2 is hardly affected at all by the new instruments. Although MSS1 does give somewhat different estimates now, they are not systematically worse than before, sometimes the relative error is better, and sometimes it is worse. It remains true that the small value for R_Z induces inefficiency. Finally, we observe that the estimated standard errors tend to be of a similar magnitude as the standard deviations, though with some exceptions. The standard errors tend to be a little higher than the standard deviations; hence, one could take standard errors as a somewhat conservative estimate of the inaccuracy of the estimates.

We conclude that for SML the limited number of drawings of 10 is clearly not sufficient for a reasonable performance of the estimator. All of the three MSS estimators perform acceptably even with only one drawing, although it is clear that efficiency can be improved by taking more than one drawing.

TABLE 2.1 MONTE CARLO RESULTS

parameter	θ_0	ML	SML $R = 10$	MSS1 $R = 1, R_Z = 10$
α_1	-8.686	-8.961	-5.996	-19.122
SD		0.586	15.689	12.004
rel. err. (%)		3.166	31.0	120.1
α_2	-0.0482	-0.0483	-0.414	-0.0503
SD		0.0146	0.653	0.0126
rel. err. (%)		0.265	760.4	4.443
α_3	3.137	3.098	-2.258	4.090
SD		0.0784	10.249	1.118
rel. err. (%)		1.259	172.0	30.4
α_4	-0.163	-0.156	-0.109	-0.210
SD		0.00850	0.0598	0.0587
rel. err. (%)		4.518	33.2	29.0
σ_ϵ^2	346.921	346.923	254.785	601.509
SD		0.0233	255.331	581.318
rel. err. (%)		0.00552	26.6	73.4
σ_u^2	1.080	1.175	0.448	1.966
SD		0.123	0.572	2.311
rel. err. (%)		8.809	58.6	82.0
η_1	-11.801	-11.860	158.243	-8.983
SD		0.0597	202.316	5.900
rel. err. (%)		0.501	1440.1	23.9
η_2	8.644	8.560	5.600	7.124
SD		0.118	6.436	3.284
rel. err. (%)		0.972	35.2	17.6
η_3	-1.199	-1.189	-3.801	-0.978
SD		0.0383	7.795	0.453
rel. err. (%)		0.844	217.0	18.4

TABLE 2.1 MONTE CARLO RESULTS (continued)

parameter	MSS1 $R=10, R_Z=10$	MSS1 $R=1, R_Z=500$	MSS2 $R=1, R_Z=10$	MSS2 $R=10, R_Z=10$
α_1	−22.589	−10.765	−2.517	−6.705
SD	11.920	6.556	6.880	3.407
rel. err. (%)	160.1	23.9	71.0	22.8
α_2	−0.0436	−0.0490	−0.0493	−0.0503
SD	0.0185	0.0136	0.0110	0.0116
rel. err. (%)	9.451	1.752	2.405	4.388
α_3	4.535	3.353	2.624	2.921
SD	1.220	0.685	0.528	0.303
rel. err. (%)	44.6	6.890	16.4	6.890
α_4	−0.221	−0.173	−0.147	−0.156
SD	0.0548	0.0300	0.0194	0.0139
rel. err. (%)	35.6	5.976	9.913	4.340
σ_ϵ^2	366.357	349.480	347.272	347.013
SD	63.981	8.146	0.469	0.163
rel. err. (%)	5.602	0.736	0.101	0.0266
σ_u^2	7.586	2.809	0.995	1.004
SD	16.004	6.503	0.0869	0.0686
rel. err. (%)	602.4	160.1	7.896	7.066
η_1	−8.430	−8.694	−12.602	−12.022
SD	3.342	5.512	5.382	0.630
rel. err. (%)	28.6	26.3	6.783	1.871
η_2	7.299	7.100	9.215	8.859
SD	1.683	2.963	2.970	0.389
rel. err. (%)	15.6	17.9	6.609	2.482
η_3	−1.003	−0.990	−1.291	−1.243
SD	0.227	0.423	0.411	0.0689
rel. err. (%)	16.4	17.4	7.707	3.658

TABLE 2.1 MONTE CARLO RESULTS (continued)

parameter	MSS3 $R = 1$	MSS3 $R = 10$
α_1	-11.786	-8.690
SD	8.002	0.0136
rel. err. (%)	35.7	0.00420
α_2	-0.0445	-0.0506
SD	0.0193	0.0141
rel. err. (%)	7.609	5.125
α_3	3.305	3.091
SD	0.641	0.178
rel. err. (%)	5.355	1.460
α_4	-0.168	-0.161
SD	0.0230	0.0100
rel. err. (%)	3.233	0.982
σ_ϵ^2	342.439	346.921
SD	19.954	0.000762
rel. err. (%)	1.292	0.0475×10^{-3}
σ_u^2	1.116	1.065
SD	0.155	0.0994
rel. err. (%)	3.332	1.427
η_1	-12.031	-11.775
SD	7.450	0.0955
rel. err. (%)	1.946	0.225
η_2	8.651	8.696
SD	4.092	0.167
rel. err. (%)	0.0850	0.599
η_3	-1.185	-1.216
SD	0.565	0.0544
rel. err. (%)	1.127	1.418

TABLE 2.2 ADDITIONAL MONTE CARLO RESULTS, TWO STAGES

parameter	MSS1 $R = 10, R_Z = 10$	MSS2 $R = 10, R_Z = 10$
α_1	-19.261	-6.146
SD	11.771	4.086
Est. SE.	14.623	9.369
rel. err. (%)	121.8	29.242
α_2	-0.0560	-0.0508
SD	0.0177	0.0121
Est. SE.	0.0106	0.00989
rel. err. (%)	16.208	5.560
α_3	4.068	2.902
SD	1.073	0.366
Est. SE.	1.318	0.607
rel. err. (%)	29.678	7.490
α_4	-0.211	-0.156
SD	0.0528	0.0154
Est. SE.	0.0674	0.0231
rel. err. (%)	29.155	4.575
σ_ϵ^2	647.011	347.050
SD	460.838	0.210
Est. SE.	493.617	103.994
rel. err. (%)	86.501	0.371×10^{-3}
σ_u	5.095	1.0001
SD	12.633	0.0756
Est. SE.	2.673	0.0696
rel. err. (%)	371.8	7.335
η_1	-11.286	-11.650
SD	12.470	10.542
Est. SE.	14.975	11.945
rel. err. (%)	4.371	0.152
η_2	8.814	8.647
SD	7.301	4.928
Est. SE.	8.539	6.734
rel. err. (%)	1.970	0.299×10^{-3}
η_3	-1.224	-1.212
SD	1.020	0.843
Est. SE.	1.191	0.951
rel. err. (%)	2.119	1.098

2.4.2 Estimation results

We now present the estimates of the model for the real data. The model has been estimated using the three methods described in section 2.3 and by ML. For each method, the model was estimated under the assumption $\sigma_{\epsilon u} = 0$ with two different numbers of drawings; R = 5 and R = 10. The restriction $\sigma_{\epsilon u} = 0$ is relaxed below. The instrument matrices are constructed using $R_Z = 10$ drawings.

In table 2.3 the ML estimates are given. The estimates have the expected sign. The maximum number of hours worked occurs at a wage rate of about 11 guilders per hour; the wage rate reaches its maximum at the age of 35 years.

In table 2.4 the estimates by method 1, using R = 5 drawings from the standard normal distribution, and their asymptotic standard errors are presented. To circumvent problems with consistency, we used the estimates of method 3 (see table 2.9) to construct the matrix of instruments. Most of the parameter estimates have the expected sign. Non-labour income has a negative impact on labour supply. The estimate of the linear wage parameter is positive and the estimate of the squared wage parameter is negative. We can calculate that the number of hours worked reaches its maximum at a wage rate of Dfl. 13.5 per hour. Log-family size influences the log-wage rate negatively. The education dummies have the expected negative sign. Moreover, the higher the level of education, the less negative is the parameter estimate of the corresponding education dummy. From the parameter estimates of log-age and its square we can calculate that the wage rate reaches it maximum at the age of 38.

The estimates in table 2.5 are also obtained by applying method 1, but now we have used R = 10 drawings to construct the simulators. Again, most of the estimates have the expected sign. The maximum of the number of hours worked with respect to the wage rate is reached at w = 13.4. The log-wage rate reaches its maximum with respect to age at 38 years. The estimates don't differ much from the ones in table 2.4, but the standard errors are somewhat lower.

In table 2.6 we present the estimates obtained by method 2. The matrix of instruments is constructed using the estimates in table 2.9. Apart from the other parameter estimates, now also the parameter estimate of $\hat{\eta}_3$ (number of children below the age of 6) has the expected sign. The wage is maximal at the age of 37 and the number of hours supplied is maximal at a wage rate of 54.7.

Table 2.7 shows the method 2 estimates when R = 10 drawings are used. Again, the matrix of instruments is constructed using the estimates in table 2.9. The wage equation is maximal at the age of 37, whereas labour supply is maximal at a wage rate of 32.9. The main difference between the method 1 estimates and the method 2 estimates are the parameter estimates of the labour supply function.

Finally, we look at the estimates obtained by method 3. In table 2.8 the results with R = 5 drawings are presented. We can make the same remarks about the signs of the estimates as in the previous cases. Wage is maximal at the age of 38 and hours supplied are maximal at the wage rate of 9.4. To obtain the results in table 2.9, we used R = 10 drawings to construct the simulators. Most of the standard errors of the estimates are lower than in table 2.8, and whenever they aren't lower, they are only slightly higher.

In table 2.10 we present the estimation results with R = 50 drawings. Comparing tables 2.4 and 2.8, we see that there is not much difference between the results with R = 10 and R = 50 drawings. Apparently R = 10 drawings are in this case sufficient to minimize the effect of the simulation residuals.

Comparing the three methods, we can say that method 1 is the cheapest in C.P.U.-time because it doesn't make use of simulators of the derivatives. Also, it appears to produce the smallest standard errors. However, it only makes sense to use this method when a consistent estimate is available to construct the matrix of instruments. Method 2 is the most expensive in C.P.U.-time.

Table 2.11 gives the results of estimating the model without the restriction of zero correlation between the disturbances by method 3, using 10 drawings to simulate the response probabilities. The estimates of the disturbances' variances and the covariance imply a correlation coefficient of $\rho = 0.082$ which is not significantly different from zero. A comparison of tables 2.11 and 2.9 reveals no big shifts in the parameter estimates.

To get some more feeling for the differences in estimates across methods, we present elasticities of hours worked and of participation with respect to wages. These have been calculated as "aggregate" elasticities in the sense that all wages in the sample have been raised by 5% and then hours and participation probabilities have been predicted for every individual in the sample. The observed changes in the sample averages of these quantities are used to compute the elasticities. The results are given in table 2.12.

Strikingly, ML gives elasticities that are much larger than those implied by the other methods. Method 1 is most similar to ML in this respect. It is hard to interpret these differences. In principle they would call for specification tests. Given the simplistic nature of the model and the illustrative purposes of the estimation we abstain from a specification search.[1]

As a final comparison of estimation methods, we present in table 2.13 the likelihood values corresponding to the estimates obtained by the various methods. We now see that MSS2 is closest to ML and MSS3 has the lowest likelihood value.

[1]Presumably the most important problem with the present model is that it assumes that anyone who wants to work can do so (if $h^* > 0$, one works a positive amount of hours). This assumption is far too strong; see for instance Blundell, Ham and Meghir (1987) or Kapteyn and Woittiez (1989).

TABLE 2.3 MAXIMUM LIKELIHOOD ESTIMATES

parameter		$\hat{\theta}^1$	standard error
α_1	(const.)	-86.696	10.860
α_2	(non-labour income)	-5.605 $\mathbf{10^{-2}}$	0.00936
α_3	(wage)	12.1126	1.312
α_4	(0.5× square of wage)	-0.574	0.0625
σ_ϵ^2		558.052	96.753
σ_u^2		0.135	0.0164
η_1	(const.)	-14.959	3.733
η_2	(log(fs))	-0.387	0.0766
η_3	(# child. < 6)	-8.765 $\mathbf{10^{-2}}$	0.0206
η_4	log(age)	10.322	2.110
η_5	(educ1)	-0.767	0.119
η_6	(educ2)	-0.771	0.117
η_7	(educ3)	-0.625	0.115
η_8	(educ4)	-0.317	0.112
η_9	(sec1)	0.266	0.410
η_{10}	(sec2)	5.611 $\mathbf{10^{-2}}$	0.0511
η_{11}	(square of log(age))	-1.452	0.295

TABLE 2.4 ESTIMATES BY METHOD 1
NUMBER OF DRAWINGS R = 5

parameter		$\hat{\theta}^1$	standard error
α_1	(const.)	-22.595	9.43
α_2	(non-labour income)	-6.194 $\mathbf{10^{-2}}$	0.02
α_3	(wage)	3.158	1.12
α_4	(0.5× square of wage)	-0.117	0.05
σ_ϵ^2		309.942	17.53
σ_u^2		6.819 $\mathbf{10^{-2}}$	0.01
η_1	(const.)	-11.855	2.43
η_2	(log(fs))	-0.118	0.06
η_3	(# child. < 6)	1.353 $\mathbf{10^{-2}}$	0.03
η_4	log(age)	8.198	1.39
η_5	(educ1)	-0.522	0.10
η_6	(educ2)	-0.454	0.10
η_7	(educ3)	-0.395	0.10
η_8	(educ4)	-0.156	0.10
η_9	(sec1)	0.163	0.04
η_{10}	(sec2)	6.796 $\mathbf{10^{-2}}$	0.04
η_{11}	(square of log(age))	-1.128	0.02

TABLE 2.4 ESTIMATES BY METHOD 1
NUMBER OF DRAWINGS R = 10

parameter		$\hat{\theta}^1$	standard error
α_1	(const.)	-22.595	9.09
α_2	(non-labour income)	-6.209 $\mathbf{10^{-2}}$	0.02
α_3	(wage)	3.162	1.08
α_4	(0.5× square of wage)	-0.118	0.05
σ_ϵ^2		309.942	17.12
σ_u^2		6.817 $\mathbf{10^{-2}}$	0.01
η_1	(const.)	-11.855	2.37
η_2	(log(fs))	-0.118	0.01
η_3	(# child. < 6)	1.373 $\mathbf{10^{-2}}$	0.03
η_4	log(age)	8.198	1.36
η_5	(educ1)	-0.521	0.10
η_6	(educ2)	-0.454	0.10
η_7	(educ3)	-0.395	0.10
η_8	(educ4)	-0.156	0.10
η_9	(sec1)	0.163	0.04
η_{10}	(sec2)	6.831 $\mathbf{10^{-2}}$	0.04
η_{11}	(square of log(age))	-1.129	0.02

TABLE 2.6 ESTIMATES BY METHOD 2
NUMBER OF DRAWINGS R = 5

parameter		$\hat{\theta}^2$	standard error
α_1	(const.)	-7.930	17.73
α_2	(non-labour income)	-4.815 $\mathbf{10^{-2}}$	0.01
α_3	(wage)	0.989	2.45
α_4	(0.5× square of wage)	-4.815 $\mathbf{10^{-3}}$	0.13
σ_ϵ^2		345.479	22.03
σ_u^2		7.911 $\mathbf{10^{-2}}$	0.02
η_1	(const.)	-12.182	2.77
η_2	(log(fs))	-0.198	0.10
η_3	(# child. < 6)	-4.607 $\mathbf{10^{-2}}$	0.05
η_4	log(age)	8.494	1.62
η_5	(educ1)	-0.579	0.12
η_6	(educ2)	-0.540	0.13
η_7	(educ3)	-0.435	0.11
η_8	(educ4)	-0.187	0.11
η_9	(sec1)	0.188	0.06
η_{10}	(sec2)	6.186 $\mathbf{10^{-2}}$	0.04
η_{11}	(square of log(age))	-1.177	0.23

TABLE 2.7 ESTIMATES BY METHOD 2
NUMBER OF DRAWINGS R = 10

parameter		$\hat{\theta}^2$	standard error
α_1	(const.)	-8.686	19.59
α_2	(non-labour income)	-4.817 $\mathbf{10^{-2}}$	0.01
α_3	(wage)	1.098	2.71
α_4	(0.5× square of wage)	-1.667 $\mathbf{10^{-2}}$	0.14
σ_ϵ^2		346.921	23.92
σ_u^2		7.924 $\mathbf{10^{-2}}$	0.02
η_1	(const.)	-12.441	2.84
η_2	(log(fs))	-0.202	0.01
η_3	(# child. < 6)	-4.507 $\mathbf{10^{-2}}$	0.05
η_4	log(age)	8.644	1.67
η_5	(educ1)	-0.580	0.12
η_6	(educ2)	-0.541	0.14
η_7	(educ3)	-0.437	0.11
η_8	(educ4)	-0.189	0.11
η_9	(sec1)	0.188	0.06
η_{10}	(sec2)	6.299 $\mathbf{10^{-2}}$	0.04
η_{11}	(square of log(age))	-1.199	0.24

TABLE 2.8 ESTIMATES BY METHOD 3
NUMBER OF DRAWINGS R = 5

parameter		$\hat{\theta}^3$	standard error
α_1	(const.)	-6.665	18.43
α_2	(non-labour income)	-6.680 $\mathbf{10^{-2}}$	0.01
α_3	(wage)	3.423	3.03
α_4	(0.5× square of wage)	-0.183	0.01
σ_ϵ^2		269.000	111.58
σ_u^2		6.959 $\mathbf{10^{-2}}$	0.10
η_1	(const.)	-11.382	2.97
η_2	(log(fs))	-0.115	0.09
η_3	(# child. < 6)	4.755 $\mathbf{10^{-2}}$	0.04
η_4	log(age)	7.902	1.73
η_5	(educ1)	-0.519	0.12
η_6	(educ2)	-0.438	0.12
η_7	(educ3)	-0.388	0.11
η_8	(educ4)	-0.148	0.11
η_9	(sec1)	0.166	0.05
η_{10}	(sec2)	7.139 $\mathbf{10^{-2}}$	0.04
η_{11}	(square of log(age))	-1.084	0.02

TABLE 2.9 ESTIMATES BY METHOD 3
NUMBER OF DRAWINGS R = 10

parameter		$\hat{\theta}^3$	standard error
α_1	(const.)	-7.717	13.55
α_2	(non-labour income)	-6.593 $\mathbf{10^{-2}}$	0.01
α_3	(wage)	3.137	2.35
α_4	(0.5× square of wage)	-0.163	0.12
σ_ϵ^2		308.748	125.49
σ_u^2		6.918 $\mathbf{10^{-2}}$	0.01
η_1	(const.)	-11.368	2.72
η_2	(log(fs))	-0.112	0.07
η_3	(# child. < 6)	4.146 $\mathbf{10^{-2}}$	0.03
η_4	log(age)	7.895	1.57
η_5	(educ1)	-0.517	0.11
η_6	(educ2)	-0.438	0.11
η_7	(educ3)	-0.387	0.10
η_8	(educ4)	-0.148	0.11
η_9	(sec1)	0.165	0.04
η_{10}	(sec2)	7.168 $\mathbf{10^{-2}}$	0.04
η_{11}	(square of log(age))	-1.084	0.02

TABLE 2.10 ESTIMATES BY METHOD 3
NUMBER OF DRAWINGS R = 50

parameter		$\hat{\theta}^3$	standard error
α_1	(const.)	-7.716	13.69
α_2	(non-labour income)	-6.598 $\mathbf{10^{-2}}$	0.01
α_3	(wage)	3.149	2.39
α_4	(0.5× square of wage)	-0.164	0.01
σ_ϵ^2		308.748	130.46
σ_u^2		6.932 $\mathbf{10^{-2}}$	0.01
η_1	(const.)	-11.369	2.76
η_2	(log(fs))	-0.111	0.08
η_3	(# child. < 6)	4.324 $\mathbf{10^{-2}}$	0.04
η_4	log(age)	7.894	1.59
η_5	(educ1)	-0.517	0.11
η_6	(educ2)	-0.437	0.11
η_7	(educ3)	-0.386	0.10
η_8	(educ4)	-0.147	0.11
η_9	(sec1)	0.168	0.05
η_{10}	(sec2)	7.263 $\mathbf{10^{-2}}$	0.04
η_{11}	(square of log(age))	-1.084	0.02

TABLE 2.11 ESTIMATES BY METHOD 3
NUMBER OF DRAWINGS R = 10
CORRELATED DISTURBANCES

parameter		$\hat{\theta}^3$	standard error
α_1	(const.)	-8.994	10.872
α_2	(non-labour income)	-6.592 $\mathbf{10^{-2}}$	0.01
α_3	(wage)	3.298	1.84
α_4	(0.5× square of wage)	-0.172	0.10
σ_ϵ^2		308.869	91.15
σ_u^2		7.054 $\mathbf{10^{-2}}$	0.01
$\sigma_{\epsilon u}$		0.147	0.58
η_1	(const.)	-11.728	2.75
η_2	(log(fs))	-0.115	0.07
η_3	(# child. < 6)	3.992 $\mathbf{10^{-2}}$	0.04
η_4	log(age)	8.093	1.57
η_5	(educ1)	-0.522	0.11
η_6	(educ2)	-0.442	0.11
η_7	(educ3)	-0.390	0.10
η_8	(educ4)	-0.149	0.11
η_9	(sec1)	0.169	0.05
η_{10}	(sec2)	7.060 $\mathbf{10^{-2}}$	0.04
η_{11}	(square of log(age))	-1.111	0.22

TABLE 2.12 WAGE ELASTICITIES ACCORDING
TO DIFFERENT ESTIMATION METHODS, R=10

	method 1	method 2	method 3	ML
hours	1.119	0.637	0.518	1.943
participation	0.606	0.133	0.236	1.369

2.5 Conclusions

The main purpose of this chapter has been to investigate the usefullness of MSS estimators in mixed discrete-continuous models, with a focus on the kind of model typically encountered in the analysis of labour supply. The experience in this chapter appears to be that the estimators perform quite well. In the example considered, the MSS estimators do a little worse than ML, but in more complex situations ML would simply be infeasible. This is not only a matter of computing time, but also due to the fact that in certain situations it is impossible to write down analytically the probability of certain events, whereas the events can still be simulated. Estimation by simulation techniques then turns out to be a useful tool in the analysis of labour supply models, in the sense that these techniques enable us to estimate models which cannot, or only with great difficulty, be estimated with conventional methods like maximum likelihood or the method of moments.

The simulated scores methods presented in this chapter perform satisfactory, even with a limited number of drawings. For the use of a limited number of drawings a price has to be paid in the form of a loss in efficiency, but this loss is modest. This is in stark contrast with the method of simulated maximum likelihood with a limited number of drawings, which performs poorly.

The method of simulated scores will be more useful, the higher the dimension of integration in the evaluation of response probabilities in the likelihood function is. In this context one may think of models of family labour supply with various sources of randomness. In this chapter we have only used smooth simulators. In more complex models the use of frequency simulators cannot always be avoided. Their main drawback is their discontinuity in the parameters as a result of which conventional gradient based optimization procedures cannot be used. The downhill simplex method, employed in chapter 3 for instance, is quite time consuming. This disadvantage, however, is mitigated by the possibility to use a limited number of drawings in the method of simulated scores.

Chapter 3

The estimation of utility consistent labour supply models by means of simulated scores

3.1 Introduction

By now there is an enormous literature on the estimation of static models of individual labour supply. Typically, a model will consist of two equations, a wage equation and a labour supply equation. Especially since the work of Hausman (1979, 1980, 1985) the labour supply equation is usually utility consistent[1] and often the underlying budget set is piecewise linear and possibly non-convex. See e.g. Blomquist (1983), Moffit (1986) and the papers in the special issue of the *Journal of Human Resources*, Summer 1990.

Despite the vast quantity of papers written on the topic, there are still various unsatisfactory elements in the models estimated so far. These pertain to both the specification and the estimation of the models. As to specification, one usually adheres to simple forms of the labour supply function, whereas the stochastic specification is often more governed by considerations of convenience than of plausibility. Estimation of simple models is not much of a problem (e.g. of a type II Tobit model), but in somewhat more complicated models often short cuts are being taken that strictly speaking impair consistency of estimators. In section 3.2 these issues will be discussed in more detail. There are good reasons for all of this. As we will illustrate in section 3.2, essentially the canonical Hausman model with a flexible specification of preferences, a non-convex budget set and a proper stochastic specification could not be estimated with methods available until recently. Possibly the most glaring difficulty is that except in very simple models it is impossible to write down the probability of participation. Since this probability plays a role in any estimation method one would like to apply, ranging from ML to MM, all estimation methods applied in practice can be seen as approximations with varying degrees of accuracy.[2]

[1]By "utility consistent" we mean throughout this chapter that observed or predicted labour supply can be rationalized as the result of the maximization of a well- behaved utility function; we call a utility function well- behaved if it is strictly quasi-concave and increasing in consumption

[2]In this chapter we do not pay attention to more extensive models where the hours decision and

Rather closely related to the previous issues is the issue of coherency. [3] It turns out that in models with kinked budget constraints coherency requires quasiconcavity of the direct utility function at all kink points. Since these kink points will be different for different individuals in a sample, coherency requires quasiconcavity at many combinations of hours and wages. See MaCurdy, Green, Paarsch (1990) or Van Soest, Kooreman, and Kapteyn (1993). This in turn means that parameters and error distributions have to be restricted in order to make sure that the model is utility consistent (i.e., the direct utility function is strictly quasi-concave) at relevant kink points for each observation. [4] Except for simple models, the imposition of utility consistency is non-trivial. Failure to do so, however, may lead to inconsistent estimators.

In this chapter we specify a utility consistent static model of individual labour supply, with a flexible specification of preferences. The model is of the conventional two-equation form, a wage equation and a labour supply equation. Three random errors are specified, one additive random error in the wage equation representing unobserved heterogeneity, one additive error in the labour supply equation representing optimization and reporting errors and a non-additive error in the labour supply equation representing random preferences. We impose utility consistency at all data points. The estimation method is a variant of a method of simulated scores (MSS), developed in an earlier paper (Bloemen and Kapteyn, 1993a), chapter 2 of this thesis. Thereby we avoid the impossible task of writing down the probability of participation; instead we draw from the errors in the model and simply determine whether utility is maximal while working or not working. Since the model and estimation method are rather intricate, we first look at the properties of the estimation method by means of Monte Carlo. Next the model is estimated for Dutch data on married females. In recent years substantial advances have been made in the development of computationally efficient simulators. See e.g. the survey by Hajivassiliou (1991). All these approaches exploit to some extent specific properties of the model at hand, like linearity, normality, or smoothness. In the present context none of these properties applies. Hence the estimation method will be rather brute force, using frequency simulators and numerical approximations of derivatives whenever required.

The order of presentation is as follows: In the next section we set out the basic model and discuss various approaches in the literature using the model as an illustration. In section 3.3 the estimator is presented. In section 3.4 we give a detailed specification of the model and the restrictions that have to be imposed to render the model coherent. In section 3.5 we present details of the simulation methods needed to operationalize the estimator. In section 3.6 we compare our estimator with a number of alternatives used in the literature. We first do this for artificial data generated by means of Monte Carlo and next for real data. Section 3.7 concludes.

---

participation decision are modelled separately, as in e.g. Blundell and Meghir (1987) or Blundell, Ham, and Meghir (1987). Although at first sight this may seem to circumvent the problem mentioned, a fully consistent treatment will still require the computation of the probability that desired hours are zero, and that is precisely the problem we are dealing with.

[3] See e.g Gourieroux, Laffont, and Montfort (1980). A model is coherent if endogenous variables are uniquely determined by the exogenous variables and the errors.

[4] If one allows for measurement error, the construction of a likelihood requires that one integrates over all possible values of the "true" number of hours, which includes all kink points.

3.2 The economic model

Consider the following utility function, which is a special case of the utility function proposed by Hausman and Ruud (1984):

$$U(h,c) = m^* \exp\left\{\frac{\beta}{\gamma}(h-\delta-\beta m^*)\right\} \tag{3.1}$$

where

$$m^* = \frac{\gamma}{\beta^2}\left[1-\left\{1+\frac{\beta^2}{\gamma}\left[\frac{(h-\delta)^2}{\gamma}-2(c+\theta)\right]\right\}^{\frac{1}{2}}\right] \tag{3.2}$$

The variables h and c are hours worked and consumption respectively; $\beta,\gamma,\delta,\theta$ are parameters. Maximization of this utility function subject to a linear budget constraint of the form $c \le wh+\mu$ yields the following labour supply function:

$$h(w,\mu) = \delta + \mu^*\beta + w\gamma \tag{3.3}$$

with

$$\mu^* = \theta + \mu + \delta w + \frac{1}{2}\gamma w^2 \tag{3.4}$$

The cost function dual to the utility function is:

$$c(u,w,\mu) = u.\exp(\beta w) - \{\theta + \delta w + \frac{1}{2}\gamma w^2\} \tag{3.5}$$

We will assume throughout that $\gamma > 0$. It is easy to see that concavity of the cost function requires

$$\mu^* < \gamma/\beta^2, \tag{3.6}$$

An equivalent condition in (c,h) space is that the indifference curves are convex. This requires:

$$m^* < \gamma/\beta^2 \tag{3.7}$$

Notice that under this condition utility is increasing in consumption.

Using (3.2), (3.7) can be written as follows:

$$\left\{1+\frac{\beta^2}{\gamma}\left[\frac{(h-\delta)^2}{\gamma}-2(c+\theta)\right]\right\}^{\frac{1}{2}} > 0 \tag{3.8}$$

Clearly this holds true whenever the utility function is defined. The condition for the existence of the utility function can be written as:

$$\frac{(h-\delta)^2}{\gamma} - 2(c+\theta) > -\frac{\gamma}{\beta^2} \tag{3.9}$$

or,

$$c < -\theta + \frac{\gamma}{2\beta^2} + \frac{(h-\delta)^2}{2\gamma} := f(h) \tag{3.10}$$

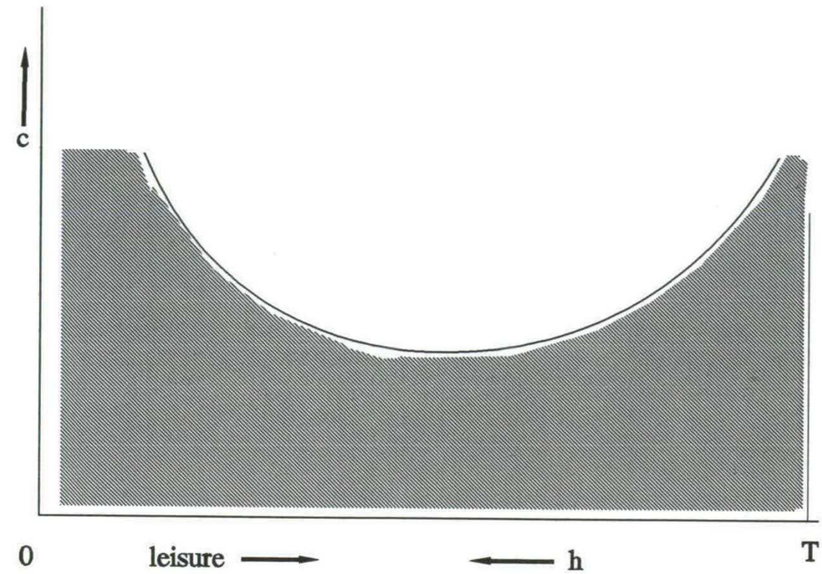

Figure 3.1: The domain of U(h,c)

with $f(h)$ implicitly defined. The function $f(h)$ is a parabola with a minimum for $h = \delta$. The value of the minimum is $-\theta + \gamma/2\beta^2$. Figure 3.1 sketches the domain of $U(h,c)$:

Let us now turn to the description of behaviour under a nonlinear (actually piecewise linear) and non-convex budget set. Figure 3.2 represents the familiar example of a utility maximum attained at the point where an indifference curve is tangent to the budget constraint.

As the budget set is not convex, but can be seen as the union of convex sets, an algorithm for finding the utility maximum is to first find points of tangency or corner solutions (kink points) for each convex set and then pick the point which yields the maximum maximorum.

To complete the model we need an equation explaining the before tax wage of an individual and the specification of the stochastic structure. Furthermore, we introduce a subscript n to index the observations, $n = 1, \ldots, N$. The wage equation is specified as follows:

$$\log w_n = \sum_{j=1}^{J} \eta_j x_{nj} + u_n \tag{3.11}$$

where u_n is an i.i.d. error term representing unobserved heterogeneity, x_{nj} are observable characteristics and η_j are parameters. As to the labour supply equation, we introduce preference variation by allowing θ to vary across agents as follows:

$$\theta_n = \theta_0 + x_n' \omega + v_n \tag{3.12}$$

where x_n is a vector of individual observable characteristics and v_n an unobservable error

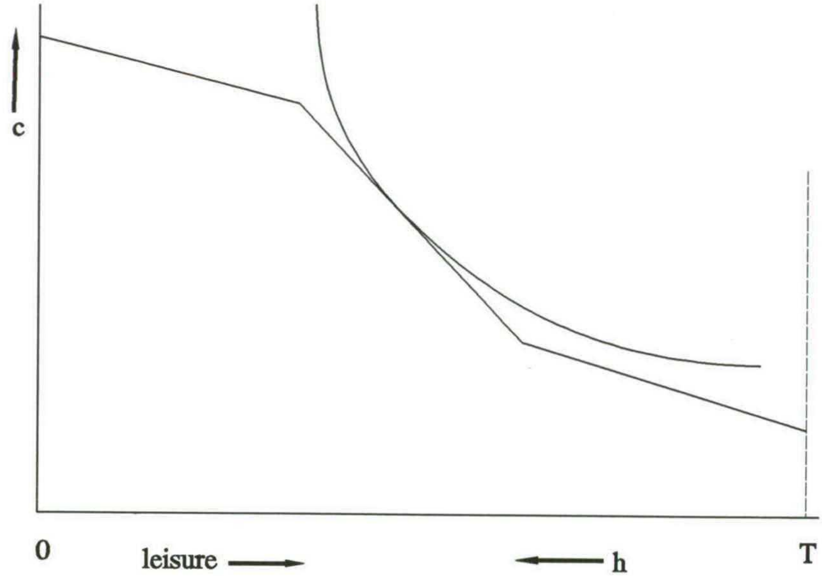

Figure 3.2: Utility maximization over a nonconvex budget set

term; ω is a parameter vector.[5] For later purposes it is useful to define also

$$\mu_n^* = \theta_n + \mu_n + w_n\delta + \frac{1}{2}w_n^2\gamma \tag{3.13}$$

For given values of w_n and v_n indivdual n's optimal number of hours, say $\overline{h}_n$ is determined as in Figure 3.2. In general this is a complicated function of the wage, nonlabour income, individual characteristics and the random preference term. We write this as

$$\overline{h}_n = h_n(w_n, \mu_n; \alpha, x_n'\omega + v_n), \tag{3.14}$$

where $\alpha = (\beta, \gamma, \delta, \theta_0)'$. Notice for instance that the function $h_n(.)$ need not be continuous. We allow for the possibility of optimization or measurement errors by adding an error term ϵ_n:

$$h_n^* = \overline{h}_n + \epsilon_n \tag{3.15}$$

Let h_n be observed labour supply, then we assume:

$$h_n = h_n^* \quad \text{if } h_n^* > 0 \quad \underline{\text{and}}\ h_n(w_n, \mu_n; \alpha, x_n'\omega + v_n) > 0 \tag{3.16}$$

$$h_n = 0 \quad \text{if } h_n^* \leq 0 \quad \underline{\text{or}}\ h_n(w_n, \mu_n; \alpha, x_n'\omega + v_n) \leq 0 \tag{3.17}$$

[5]There is no a priori reason to let preference variation enter through θ only, any of the other parameters of the utility function may be made dependent on observable and unobservable characteristics. For simplicity of the exposition we stick to the present somewhat arbitrary choice

This formulation brings out the distinction between the random preferences v_n and the optimization or measurement errors ϵ_n. This type of stochastic specification is in line with the work of Hausman (1981).

With this model at hand, we can discuss various problems in estimation and characterize different approaches in the literature.

- *It is difficult to derive the density of wages and hours for working individuals if the budget constraint is non-convex.* The joint density of hours and (before tax) wages can be written as the product of the conditional density of hours given wages and the marginal density of wages. The latter is not difficult to write down, but the former may be. As indicated above, if the budget constraint is non-convex it can be written as the union of convex sets and the obvious thing to do is to first find the utility maximum in each convex set and next compute the *maximum maximorum*. With random preferences this means that we have to find the density of hours for each convex subset and the probability that the *maximum maximorum* is in any of the convex sets. Finding the density of hours for each convex set is tedious but feasible (see section 3.5), but the probability that the utility maximum occurs in any given convex subset is almost impossible to write down, as one can easily imagine, by inspecting the formula for the direct utility function. This difficulty will arise in all but the simplest utility specifications.

- *It is very hard to write down the probability of participation.* To write down the probability of participation, one has to characterize the values of ϵ_n, u_n, v_n for which individual n will be observed working. For non-working individuals the wage they could earn while working is typically not known. The random variables u_n and v_n cause the budget constraints and the indifference curves to move around in a complicated way. Even for a given budget constraint (i.e. for someone who does participate) it is difficult to find the values of v_n for which the utility of working will exceed that of not working for the same reason as given above. Since the budget constraint is the result of the interplay of the gross wage with possibly quite complicated institutions, the resulting distribution of the budget constraint will in general be intractable. Combined with the difficulty of writing down the probability of working for a given budget constraint, this makes it impossible to write down the probability of participation as an analytic function of exogenous variables and parameters.

- *Incoherency.* The problem of finding a utility maximum is generally well-defined if indifference curves are convex. However, if a flexible specification is adopted for the utility function, it will generally not be globally quasi-concave and hence there will be combinations of the parameters and values of exogenous variables and errors for which indifference curves are not convex or are not defined, cf. Fig. 3.1. As shown by Van Soest, Kooreman, Kapteyn (1993), this means that the model is no longer coherent. This in turn implies that estimation methods are not well-defined. To have well-defined estimation, coherency has to be imposed. For instance, these authors give an example where data are generated by a coherent model, but no coherency is imposed in estimation. It is shown that in that case the "likelihood" does not attain its maximum at the true parameter point, but rather at a point which violates coherency.

- *Time consuming numerical integration.* Even if we are able to write down in principle the probability of certain events or the density of wages and hours, they are bound to be complicated expressions involving multi-dimensional integrals. Since the model involves various non-linear transformations it is very unlikely that analytical solution of the integrals is possible. Numerical integration tends to be extremely time consuming.

To solve or evade these problems, various routes can be taken.

- *Choose simple functional forms.* As said above, to write down the joint density of (before tax) wages and hours for working individuals one needs the density of hours conditional on the wage times the marginal density of the wage. The latter is straightforward. The former can be simplified considerably by choosing a simple specification, like e.g. the Hausman linear labour supply function. This function arises from model (3.1)- (3.4) by letting γ approach zero. The utility function then reduces to
$$U(h,c) = \frac{\delta - h}{\beta} \exp\left\{-1 + \frac{\beta(c+\theta)}{h-\delta}\right\} \tag{3.18}$$

- *Ignore random preferences.* This can take two forms. One can ignore random preferences altogether, so that the only source of random variation of hours given wages is optimization or reporting errors (see e.g. Kooreman and Kapteyn (1986)). Alternatively, sometimes a random error is appended to the (non-stochastic) utility difference between working and not working, as in Kapteyn, Kooreman and Van Soest(1990). This latter term may also have the interpretation of random preferences. It is of course a bit hard to see why preferences would only be random in utility comparisons and not in the hours choice. [6]

- *Ignore unobserved heterogeneity in wages for non-participants.* In this approach the wage equation is estimated for working individuals and then used to predict before tax wages for non-participants. The implied budget constraint is assumed to be the true budget constraint, with neglect of unobserved heterogeneity. Notice that for the estimation of the wage equation for working individuals some correction for selectivity bias is required, which in turn requires the probability of participation. Strictly speaking, one should use the full model to estimate this probability. Since, as indicated, this is either very difficult or impossible, some approximation of this probability is used at this stage.

- *Use working individuals only, with correction for selectivity.* By only using working individuals one does not really avoid the necessity of computing the probability of participation, but one can approximate this probability as in the previous approach. Often, if only working individuals are used in the analysis, the budget constraint is

[6] An alternative interpretation may be that the random term in the utility comparison represents optimization errors, which may be more natural in certain contexts.

linearized in the observed point. Of course the marginal wage used to linearize the budget constraint is endogenous, but this may be solved by the use of instrumental variables. In this approach, typically no steps are taken to guarantee coherency of the model in all data points. For this reason. it is not quite clear whether the estimation method is consistent or not.

3.3 Estimation

In this section one of the estimation methods described in chapter 2 is used. For ease of notation we introduce the dummy variable d_n with

$$d_n = 1 \text{ if } h_n = 0 \tag{3.19}$$

$$d_n = 0 \text{ if } h_n > 0 \tag{3.20}$$

Furthermore we write $P_n(\vartheta)$, where ϑ contains the parameters of α, η and the parameters of the distribution functions of u_n, v_n, and ϵ_n, for the probability that h_n equals zero. The joint density of wages and hours for a participating agent n is denoted by $g^*(h_n^*, w_n|x_n, \mu_n, \vartheta)$, where

$$-\infty < h_n^* < \infty$$
$$0 < w_n < \infty$$

From this we can derive the mixed discrete-continuous probability density function of h_n and w_n, $g(h_n, w_n|x_n, \mu_n, \vartheta)$.

$$g(h_n, w_n|x_n, \mu_n, \vartheta) = \begin{cases} P_n(\vartheta) & \text{if } h_n = 0 \\ g^*(h_n, w_n|x_n, \mu_n, \vartheta) & \text{if } h_n > 0, 0 < w_n < \infty \end{cases}$$

For ease of notation we will often denote the probability $P_n(\vartheta)$ by P_n. We shall denote the probability of working $1 - P_n(\vartheta)$ by $\bar{P}_n(\vartheta)$ or simply by $\bar{P}_n$. We assume that our sample is ordered in such a way that the observations 1 to N_1 refer to non-working individuals and the observations $N_1 + 1$ to N are working individuals.

Generally, the log-likelihood function of the model is

$$L(\vartheta|x_n, \mu_n, w_n, h_n, n = 1, ..., N) =$$
$$\sum_{n=1}^{N_1} \ln P_n(\vartheta) + \sum_{n=N_1+1}^{N} \ln g^*(h_n, w_n|x_n, \mu_n, \vartheta) \tag{3.21}$$

It will be assumed that the likelihood is differentiable almost everywhere with respect to the elements of ϑ[7]. Thus in principle we can differentiate the log-likelihood with respect to ϑ to derive the first order conditions for a maximum.

$$\frac{\partial L(\vartheta)}{\partial \vartheta} = \sum_{n=1}^{N_1} \frac{\partial \ln P_n(\vartheta)}{\partial \vartheta} + \sum_{n=N_1+1}^{N} \frac{\partial \ln g^*(h_n, w_n|x_n, \mu_n, \vartheta)}{\partial \vartheta} \tag{3.22}$$

$$\frac{\partial L(\hat{\vartheta}_{ML})}{\partial \vartheta} = 0 \tag{3.23}$$

[7]For consistency some additional regularity conditions are required: $\frac{\partial^2 L}{\partial \vartheta \partial \vartheta'}$ exists in a neighbourhood of ϑ_0 and is non-singular and negative definite in a neighbourhood of ϑ_0.

where $\hat{\vartheta}_{ML}$ is the maximum likelihood estimator of ϑ.

Alternatively, we can rewrite the derivative of the log-likelihood function as

$$\frac{\partial L(\vartheta)}{\partial \vartheta} = \sum_{n=1}^{N} \left[d_n \frac{\partial \ln P_n(\vartheta)}{\partial \vartheta} + (1-d_n) \frac{\partial \ln g^*(h_n, w_n | x_n, \mu_n, \vartheta)}{\partial \vartheta} \right] \tag{3.24}$$

where d_n is the dummy variable introduced above. Let ϑ_0 be the true parameter value. It is well known that if the support of h_n and w_n does not depend on ϑ, the score vector has expectation zero:

$$E\left(\frac{\partial L(\vartheta_0)}{\partial \vartheta} \right) = 0 \tag{3.25}$$

It is this fact which implies consistency of the ML estimator. In the present context the evaluation of the score vector is impossible for the reasons set out in the previous section. We will replace the score by an unbiased simulator, which can then still be used for consistent estimation of the parameters in ϑ.

We rewrite the first order derivative of the log-likelihood function in the following way.

$$\frac{\partial L(\vartheta)}{\partial \vartheta} = \sum_{n=1}^{N} \left\{ Z_n(d_n - P_n) + (1-d_n) \left[\frac{\partial \ln g^*(h_n, w_n | x_n, \mu_n, \vartheta)}{\partial \vartheta} - \frac{\partial \ln \bar{P}_n}{\partial \vartheta} \right] \right\} \tag{3.26}$$

where

$$Z_n := \frac{\frac{\partial P_n}{\partial \vartheta}}{P_n(1-P_n)} \tag{3.27}$$

The first component of this expression equals the score of the log-likelihood of the binary response model. If we replace the vector Z_n by an arbitrary vector of instruments $\bar{Z}_n$, independent of ϑ, the expectation of the resulting expression, conditional on $\bar{Z}_n$, equals zero at the true parameter value ϑ_0.

In Section 3.5 we will describe how we simulate $P_n(\vartheta)$ and its derivative unbiasedly. As to the second term in (3.26) the simulation of the part involving the density g^* will be described in Section 3.5 as well. And finally, we use the fact that

$$E\left[(1-d_n) \frac{\partial \ln \bar{P}_n}{\partial \vartheta} \right] = \frac{\partial \bar{P}_n}{\partial \vartheta} \tag{3.28}$$

so that we may replace $(1-d_n)\frac{\partial \ln \bar{P}_n}{\partial \vartheta}$ by $\frac{\partial \bar{P}_n}{\partial \vartheta}$ without affecting the unbiasedness property of the score. As a result, the original score vector is replaced by

$$\frac{\partial \bar{L}}{\partial \vartheta} = \sum_{n=1}^{N} \left[\bar{Z}_n(d_n - P_n) + (1-d_n) \frac{\partial \ln g^*(h_n, w_n | x_n, \mu_n, \vartheta)}{\partial \vartheta} - \frac{\partial \bar{P}_n}{\partial \vartheta} \right] \tag{3.29}$$

Inserting unbiased simulators k_{nR} and $\bar{m}_{nR}$ based on R replications for the response probabilities and their derivatives respectively in this expression gives the *simulated score*:

$$K_R(\vartheta) = \sum_{n=1}^{N} \left[\bar{Z}_n(d_n - k_n(\vartheta, v_R^*)) + (1-d_n) \frac{\partial \ln g^*(h_n, w_n | x_n, \mu_n, \vartheta)}{\partial \vartheta} - \bar{m}_n(\vartheta, v_R^*) \right] \tag{3.30}$$

The advantage of writing the score vector this way is that the simulators for the response probabilities and their derivatives enter the expression linearly. As a result simulation errors are averaged out over individuals. Moreover, if a frequency simulator is used to simulate the response probabilities discontinuities are averaged over individuals, thereby eliminating the reason for the poor performance of the frequency simulator in the context of simulated maximum likelihood estimation. See, e.g., Lerman and Manski (1981) and Börsch-Supan and Hajivasiliou (1993).

The estimation procedure now becomes: Choose instrument vectors $\bar{Z}_n$ and obtain the estimator by solving the moment conditions:

$$K_R(\vartheta) = 0 \tag{3.31}$$

To ascertain the efficiency of the estimator described here, we compare it to the ML estimator. A convenient way of doing this is to look at the "simulation residual", i.e. the difference between the score and the simulated score. First, (3.26) with Z_n replaced by $\bar{Z}_n$ is compared with (3.30). Then the following residual is obtained:

$$\frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R}\left[\bar{Z}_n(P_n - k_{nr}) - \left\{\bar{m}_{nr} - (1-d_n)\frac{\partial \ln \bar{P}_n}{\partial \vartheta}\right\}\right] \tag{3.32}$$

The dummy variable can be rewritten as

$$\begin{aligned} d_n &= P_n + \nu_n \\ \text{with } E(\nu_n) &= 0 \\ \text{and } Var(\nu_n) &= P_n(1-P_n) \end{aligned} \tag{3.33}$$

Inserting this in the residual gives:

$$\frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R}\left[\bar{Z}_n(P_n - k_{nr}) - \left\{\bar{m}_{nr} - \frac{\partial \bar{P}_n}{\partial \vartheta}\right\}\right] - \sum_{n=1}^{N}\nu_n\frac{\partial \ln \bar{P}_N}{\partial \vartheta} \tag{3.34}$$

The variance of the first term of (3.34) can be reduced by increasing the number of drawings R. Suppose that $Var[\bar{Z}_n(P_n - k_{nr}) - \{\bar{m}_{nr} - \frac{\partial \bar{P}_n}{\partial \vartheta}\}] = \Xi_n$, conditional on the instruments $\bar{Z}_n$. This variance does not depend on R because the drawings are identical and independent. Then the variance of the first term is $\frac{1}{R}\sum_{n=1}^{N}\Xi_n$. With fixed N, increasing R to infinity results in reducing this variance to zero. The second term is the error which is caused by the fact that $(1-d_n)\frac{\partial \ln \bar{P}_n}{\partial \vartheta}$ is simulated by a simulator for $\frac{\partial \bar{P}_n}{\partial \vartheta}$. The expectation of this term equals zero, whereas the variance equals $\sum_{n=1}^{N}\bar{P}_n P_n \frac{\partial \ln \bar{P}_n}{\partial \vartheta}\frac{\partial \ln \bar{P}_n}{\partial \vartheta'}$. This term of the simulation residual cannot be influenced by the number of drawings. Therefore, this term leads to inefficiency, also for large R.

To compare the efficiency of the method of simulated scores estimator to the maximum likelihood estimator, the term involving the difference between $\bar{Z}_n$ and Z_n also has to be taken into account. The simulation residual then becomes:

$$RES = \frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R}\left[Z_n(P_n - k_{nr}) - \left\{\bar{m}_{nr} - \frac{\partial \bar{P}_n}{\partial \vartheta}\right\}\right] + \sum_{n=1}^{N}(\bar{Z}_n - Z_n)(d_n - k_n) - \sum_{n=1}^{N}\nu_n\frac{\partial \ln \bar{P}_n}{\partial \vartheta} \tag{3.35}$$

Compared to (3.34) we now have an additional term involving the difference $\bar{Z}_n - Z_n$. If we base $\bar{Z}_n$ on (3.27), with some consistent estimate of ϑ and simulators of $\frac{\partial \bar{P}_n}{\partial \vartheta}$ and P_n based on R_Z replications, then for R_Z going to infinity the difference disappears. Note that during the optimization process the matrix of instruments $\bar{Z}$ is fixed. This implies that this matrix only needs to be initialized once at the beginning of the estimation procedure. Therefore it is computationally feasible to calculate the instrument matrix using a large number of drawings R_Z, which need not be equal to the number of drawings R that is used in the calculation of the remaining part of the simulated score.

3.3.1 Asymptotic distribution of the estimator

The simulated score vector satisfies the property that its expectation, evaluated at the true parameter vector ϑ_0, equals zero. It is intuitively clear that if we solve the moment equations, defined by the simulated score, with respect to the parameter vector, the resulting parameter vector $\hat{\vartheta}_R$, at which the simulated score is zero, will converge to the true parameter value ϑ_0, or, equivalently, $\hat{\vartheta}_R$ will be a consistent estimator of ϑ_0. If smooth simulators were used, standard asymptotic theory could be applied to derive the consistency and asymptotic normality of the estimator. Pakes and Pollard (1989) derive conditions for consistency and asymptotic normality that do not rely on smoothness assumptions. The first set of conditions concerns the moment vector (3.29). Let $G(\vartheta)$ denote the unconditional expectation of the term in square brackets in (3.29). $G(\vartheta)$ is a set of population moments. Note that $G(\vartheta_0) = 0$. The identifiability condition requires that $\inf_{\|\vartheta - \vartheta_0\| > \delta} \| G(\vartheta) \| > 0 \forall \delta > 0$. Furthermore, $G(\vartheta)$ is required to have a non-singular derivative matrix at ϑ_0.

$\frac{1}{N} K_R(\vartheta)$ is the empirical counterpart of $G(\vartheta)$. In their proof of consistency and asymptotic normality, Pakes and Pollard use the fact that $\frac{1}{N} K_R(\vartheta)$ can be written as the expectation with respect to an empirical distribution function, the shape of which depends on the particular simulator that is employed, which should converge to its population counterpart $G(\vartheta)$, making use of the independence-across-observations assumption. Consequently, no smoothness assumptions are required. A condition which has to be satisfied in case of using a frequency simulator is that the probability of being at a tie (i.e. $d = 1$ and $d = 0$) has to be zero at ϑ_0. This condition is clearly satisfied here. Finally, Pakes and Pollard, as opposed to McFadden, allow the region which determines whether $d = 1$ or $d = 0$ to be a non-smooth function of the parameters as well. For our application non-smoothness of this region is not required. Smoothness of this region, together with the above conditions, is sufficient for consistency and asymptotic normality:

$$\frac{1}{\sqrt{N}} K_R(\vartheta_0) \,\widetilde{\text{asy}}\, N(0, V_{R0}), \tag{3.36}$$

with V_{R0} some positive definite symmetric matrix, and

$$\sqrt{N}(\hat{\vartheta}_R - \vartheta_0) \quad \widetilde{\text{asy}} \quad N(0, \Gamma^{-1} V_{R0} (\Gamma')^{-1}) \tag{3.37}$$

$$\text{where } \Gamma' \quad = \quad plim \frac{1}{N} \frac{\partial(\frac{\partial L(\vartheta_0)}{\partial \vartheta'})}{\partial \vartheta} \tag{3.38}$$

Using the expression of the asymptotic covariance matrix and the results of the analysis of the simulation residuals, it is possible to analyse the efficiency of the estimators by comparing the asymptotic covariance matrices of the simulation estimators with the asymptotic covariance matrix of the maximum likelihood estimator. It is a well known result that

$$\sqrt{N}(\hat{\vartheta}_{ML} - \vartheta_0) \quad \widetilde{\text{asy}} \quad N(0, \Omega_{ML}) \tag{3.39}$$

$$\text{where } \Omega_{ML} \quad = \quad B^{-1} \tag{3.40}$$

$$B = -plim \frac{1}{N} \frac{\partial^2 L(\vartheta_0)}{\partial \vartheta \partial \vartheta'} \tag{3.41}$$

To make clear the relation with the asymptotic covariance matrix of the simulation estimators we rewrite Ω_{ML} as

$$\Omega_{ML} = \Gamma_{ML}^{-1} V_{ML} (\Gamma'_{ML})^{-1} \tag{3.42}$$

where

$$\Gamma'_{ML} = plim \frac{1}{N} \frac{\partial \left(\frac{\partial L(\vartheta_0)}{\partial \vartheta'} \right)}{\partial \vartheta} = -B \tag{3.43}$$

which is the equivalent of (3.40), and

$$V_{ML} = plim \frac{1}{N} \sum_{n=1}^{N} \frac{\partial L_n(\vartheta_0)}{\partial \vartheta} \frac{\partial L_n(\vartheta_0)}{\partial \vartheta'} \tag{3.44}$$

To examine the efficiency of the estimator we first need to establish the relation between Γ'_{ML} and Γ'. It is readily established that

$$\Gamma' - \Gamma'_{ML} =$$
$$-plim \tfrac{1}{N} \textstyle\sum_{n=1}^{N} \left[(\bar{Z}_n - Z_n) \frac{\partial P_n}{\partial \theta} + \frac{\partial Z'_n}{\partial \theta}(d_n - P_n) + \nu_n \frac{\partial \left(\frac{\partial \ln \bar{P}_n}{\partial \theta'} \right)}{\partial \theta} + \frac{1}{\bar{P}} \frac{\partial \bar{P}_n}{\partial \theta} \frac{\partial \bar{P}_n}{\partial \theta'} \right] \tag{3.45}$$

from which only the first three terms equal zero if the instruments are constructed according to (3.27) with the number of drawings per individual tending to infinity. From the analysis of the simulation residuals it becomes clear that if the matrix of instruments is constructed according to (3.27) with drawings tending to infinity, and if the response probabilities and their derivatives are simulated with R tending to infinity as well, the asymptotic variance of the score of the likelihood function, evaluated in a consistent estimator is exceeded by X, where

$$X = plim \left(\frac{1}{N} \sum_{n=1}^{N} P_n \bar{P}_n \frac{\partial \ln \bar{P}_n}{\partial \vartheta} \frac{\partial \ln \bar{P}_n}{\partial \vartheta'} \right) \tag{3.46}$$

To estimate the covariance matrix we calculate

$$\hat{\Omega}_R = \hat{\Gamma}^{-1} \hat{V}_R (\hat{\Gamma}')^{-1} \tag{3.47}$$

with

$$\hat{\Gamma}' = \frac{1}{N}\frac{\partial(\frac{\partial \bar{L}(\hat{\vartheta}_R)}{\partial \vartheta'})}{\partial \vartheta} \tag{3.48}$$

$$\hat{V}_R = \frac{1}{N}\sum_{n=1}^{N} K_{nR}(\hat{\vartheta}_R)K_{nR}(\hat{\vartheta}_R)' \tag{3.49}$$

where the index n indicates the n-th component of the simulated score. Expression (3.49) can be calculated by simulation.

3.4 Stochastic specification

Recall (3.12):

$$\theta_n = \theta_0 + x_n'\omega + v_n \tag{3.50}$$

We have seen above that for the specification of the utility function adopted here, indifference curves will be convex whenever the utility function is defined. The utility function is defined whenever (3.10) holds true. We want (3.10) to hold true for all data points. To indicate this, we add subscripts and write:

$$c_n < -\theta_n + \frac{\gamma}{2\beta^2} + \frac{(h_n - \delta)^2}{2\gamma} := f_n(h_n) \tag{3.51}$$

To ensure that the direct utility function is properly defined for every individual in the sample, a practical procedure is the following one. Let $\tilde{w}$ and $\tilde{\mu}$ be the wage rate and non-labour income which imply a linear budget constraint such that all observed budget sets are contained in it. We call this an *encompassing budget set.* See Fig. 3.3 for an illustration.

If we restrict the range of θ_n such that inequality (3.51) holds for all values of c_n and h_n in this encompassing budget set, then we know that indifference curves are convex at all data points. To achieve this we have to restrict the range of θ_n such that the function $f_n(.)$ is either tangent to the encompassing budget constraint or is outside the encompassing budget set. A tangency point is found for $h_n = \delta + \tilde{w}\gamma$ and

$$\theta_n = -\tilde{\mu} - \delta\tilde{w} - \frac{1}{2}\gamma\tilde{w}^2 + \frac{\gamma}{2\beta^2} \tag{3.52}$$

Thus, in view of (3.51) the inequality constraint on θ_n has to be

$$\theta_n < -\tilde{\mu} - \delta\tilde{w} - \frac{1}{2}\gamma\tilde{w}^2 + \frac{\gamma}{2\beta^2} \tag{3.53}$$

To guarantee that this inequality on θ_n holds for all observations we proceed as follows. Let the error term v_n be defined on $(-\infty, 0)$ then we impose the restriction:

$$\theta_0 < -x_n'\omega - \tilde{\mu} - \delta\tilde{w} - \frac{1}{2}\gamma\tilde{w}^2 + \frac{\gamma}{2\beta^2} \tag{3.54}$$

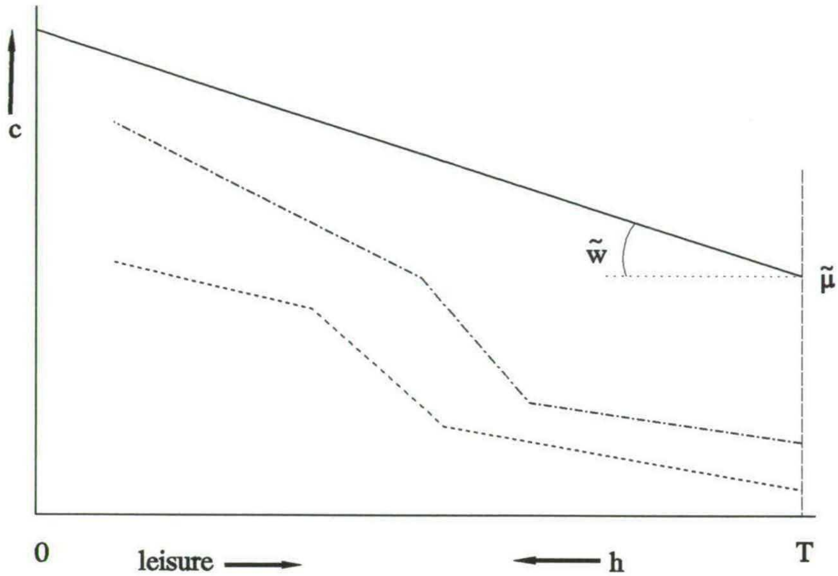

Figure 3.3: An encompassing budget set

for all n. For the random preference term v_n we will actually assume that it follows a negative Γ distribution, defined on $(-\infty, 0)$. A similar procedure, in a somewhat different context, was followed by Kapteyn, Kooreman, and Van Soest (1990).

For non-participating individuals wages are not known and have to be integrated out. To ensure coherency of the model, the support of the wage distribution has to be restricted so that for all wages the implied budget set is contained within the encompassing budget set. This is achieved by restricting the support of the wage distribution to $[0, \tilde{w}]$. A convenient choice of distribution for u_n, which restricts the range of w_n to $[0, \tilde{w}]$, is to define a random variable λ_n following a lognormal distribution with log-mean m_n and log-variance τ^2, and to define

$$u_n := \log\{\psi_n \lambda_n / [1 + \lambda_n]\} \tag{3.55}$$

where

$$\psi_n := \tilde{w} / [\exp(w(x_n, \eta))] \tag{3.56}$$

It seems reasonable to require that the median of $\exp(u_n)$ equals one. This holds true if we specify

$$m_n = -\log(\psi_n - 1) \tag{3.57}$$

Thus the number of free parameters is equal to the case where the u_n were assumed normal with mean zero.[8]

[8]Notice that under normality the inequality (3.53) is violated with non-zero probability and hence the model would be incoherent.

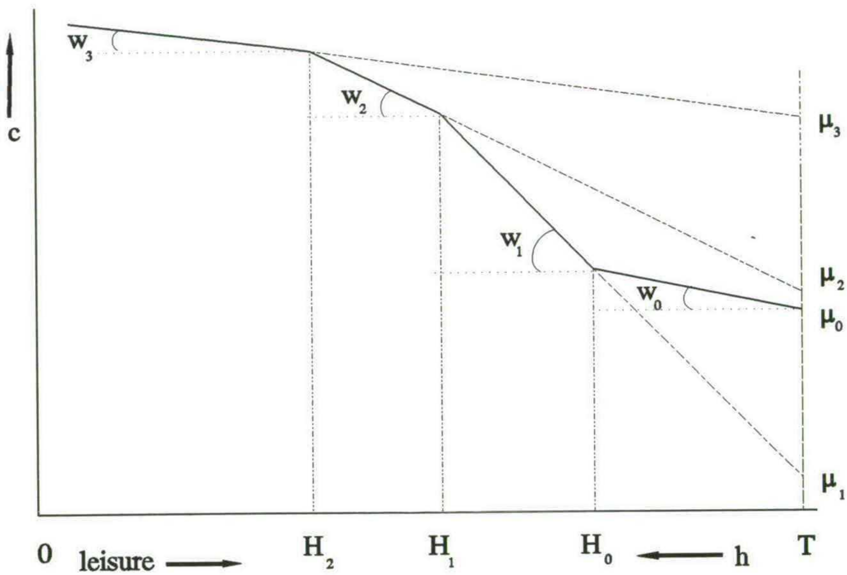

Figure 3.4: The determination of optimal hours

3.5 Outline of simulation

In this section a description is given of the construction of the simulators. The technical details are presented in the appendix. Most of the simulation can be understood by reconsidering Fig. 3.2. For convenience, we repeat the basic features in Fig. 3.4 and add some notation. Figure 3.4 presents an example of a non-linear budget constraint and a non-convexity where the non-linearities arise from a tax system with two brackets and from the welfare and social security system. The budget constraint has three segments. In the first segment, on the right hand side of the figure, the individual works a positive number of hours, whereas at the same time he receives an unemployment benefit. Of each additional guilder of labour income the individual looses, say, $\alpha\%$ of the social security benefit, until, eventually, at hours H_0, nothing is left of the benefit. This results in a net wage rate in the first segment of w_0. In the second segment, between H_0 and H_1, no more benefits are received and therefore the net wage rate rises to w_1. At hours H_1 the next tax bracket is reached, which causes the net wage rate to fall to w_2 in the third segment. Now let h_n^j denote the optimal labour supply of individual n at a linear budget constraint with slope w_{nj} and intercept μ_{nj}, $j = 0, 1, ..., m$, and denote optimal labour supply by $\bar{h}_n$.

$$\bar{h}_n = h(w_n, \mu_n; \alpha, x_n'\omega + v_n) \tag{3.58}$$

in which w_n is the before tax wage rate which implies that the function $h(.)$ includes the tax and the welfare system. Then the optimal labour supply $\bar{h}_n$, conditional on v_n and

w_n, can be determined according to the following scheme:

$$
\begin{array}{rlllllll}
h_{n,NC} & = 0 & if & & & h_n^0 & \leq & 0 \\
 & = h_n^0 & if & 0 & \leq & h_n^0 & \leq & H_{n0} \\
 & = H_{n0} & if & & & h_n^0 & > & H_{n0} \\
\tilde{h}_n & = H_{n0} & if & & & h_n^1 & < & H_{n0} \\
 & = h_n^j & if & H_{n,j-1} & < & h_n^j & \leq & H_{nj} \quad j = 1, ..., m-1 \\
 & = H_{nj} & if & h_n^j & > & H_{nj} & > & h_n^{j+1} \quad j = 1, ..., m-1 \\
 & = h_n^m & if & H_{n,m-1} & \leq & h_n^m & < & T \\
 & = T & if & & & h_n^m & \geq & T
\end{array}
\tag{3.59}
$$

$$
\begin{array}{rll}
\bar{h}_n & = \tilde{h}_n & \text{if utility in } \tilde{h}_n \text{ exceeds utility in } h_{n,NC} \\
 & = h_{n,NC} & \text{otherwise}
\end{array}
$$

$$h_n^j = \alpha_1 + \alpha_2 \mu_{nj} + \alpha_3 w_{nj} + \frac{1}{2} \alpha_4 w_{nj}^2 + x_n' \zeta + \alpha_2 v_n$$

The parameters α and ζ are obtained by reparametrization of β, γ, δ and ω in section 3.2. The precise form of this reparametrization is given in the appendix. The utility level has to be calculated using the direct utility function. Note, that if the coherency restrictions are satisfied, the event $\bar{h}_n = H_{n0}$ will occur with probability zero.

Looking at variation in the gross wage rate w_n and in the unobservable taste component v_n, we can say that w_n determines the segments of the budget constraint by determining the slopes and the kink points, whereas v_n determines in which segment labour supply will be optimal given everything else. The distribution of measurement or optimization errors ϵ_n can be used to determine the distribution of h_n^*, defined in (3.15), conditional on the unobserved taste component v_n and on the gross wage rate w_n. The distribution of v_n can be used to integrate out the unobserved taste component, taking into account the decision rule on $\bar{h}_n$. Note that this involves comparing and integrating over utility functions. Finally, we multiply by the marginal density of the wage rate to obtain the joint density of h_n^* and w_n. This joint distribution can be used to determine the censored distribution of observed labour supply h_n and it will be clear that the expression for the probability that observed labour supply equals zero ($h_n = 0$) will be complicated. It will be difficult even to write down an analytic expression for the probability and the likelihood function as a whole. Therefore it will be impossible to use smooth simulators (see, e.g. McFadden,1989), because an analytic expression is needed in order to construct a smooth simulator. We will simulate the probability with a frequency simulator F_{nR}. The frequency simulator works as follows: Draw R times a wage rate w_{nr}^*, an unobserved taste variable v_{nr}^* and a measurement error ϵ_{nr}^* from their assumed distributions, calculate $\bar{h}_n$ and h_n^* and use the rules in (3.16) and (3.17) to determine h_n. The simulator becomes:

$$f_{nr} = 1 \quad \text{if } h_n > 0 \tag{3.60}$$

$$f_{nr} = 0 \quad \text{otherwise} \tag{3.61}$$

$$F_{nR} = \tfrac{1}{R} \textstyle\sum_{r=1}^{R} f_{nr} \tag{3.62}$$

Since we also need the vector of derivatives of $P_n(\vartheta)$ we approximate this vector by a difference approximation of frequency simulators which is an unbiased simulator for the difference approximation of the probabilities. The evaluation of the contribution to the score vector of the working individuals involves the integration over random preferences. Two different methods to implement the integration are discussed in the appendix.

Finally, a suitable algorithm has to be chosen to minimize the objective function which can handle the discontinuities caused by the use of frequencies. Methods which make use of first derivatives turned out not to work and therefore we switched to the downhill simplex method of Nelder and Mead (1965) of which an overview is given in Press et al. (1986).

3.6 Results

In this section, the model is estimated using Monte Carlo data as well as real data. In the Monte Carlo experiment, the performance of the MSS method, outlined in section 3.3, is compared with some of the more conventional methods that have been mentioned in section 3.2. The conventional methods that we consider are the estimation of the model without random preferences and the estimation by instrumental variable methods. Furthermore, two variants of the MSS method are applied. The first variant estimates the parameters of the wage distribution separately by means of a reduced form wage-participation model. Then the labour supply parameters are estimated using predicted wages for non-participants. The second variant consists of estimating the wage and labour supply parameters simultaneously, thereby taking into account the stochastic nature of the budget constraint.

Subsequently the model is estimated for a sample of 849 married females, drawn from the Dutch population in 1985. In the Monte Carlo experiment the real data series of exogenous variables has been used in conjunction with a priori chosen parameter values to generate endogenous variables for each observation. The vector of taste shifters consists of the log family size variable (parameter ζ_1) and a dummy indicator taking the value one if the woman has children with age below 6, and zero if not (parameter ζ_2). In the Monte Carlo experiment, the variables in the wage equation are a constant term (parameter η_1), log-age (parameter η_2) and log-age squared (parameter η_3). To restrict the number of parameters in the Monte Carlo experiment we have omitted dummy indicators for the level of education. These dummies are included in the estimation on the real data. The parameters of the negative gamma distribution are γ_1 and γ_2. Using the distributional assumptions, random numbers have been generated which have been transformed to hours and wages using the true parameter values in the first column of table 3.1 and the decision rules in (3.16) and (3.17). The true parameter values are chosen such that the coherency restrictions are satisfied. Moreover we have tried to choose parameter values that generate distributions of observables similar to what we see in the sample. The values of some parameters are the result of experimentation with preliminary versions of the mode. Benefits are measured in guilders per week.

Both for the real data and the Monte Carlo data the budget constraint of each individual has been constructed on the basis of the Dutch tax code, also taking into account the welfare and social security system. In 1985 the tax system and the social security system were not well-integrated. They each have their own marginal tax rates and the social security system has ceilings for different sorts of payroll taxes. As a result of this the budget constraint may be quite complex with various kinks and with non-convexities.

Table 3.1 presents the a priori chosen parameter values which are used to generate the Monte Carlo data. Twenty Monte Carlo datasets have been generated using the completely specified model, which includes random preferences. [9]

Table 3.1: True parameter vector

Labor supply equation	
α_1 (Const.)	14.14
α_2 (non-labour income)	-0.04692
α_3 (wage)	10.690
α_4 (0.5× square of wage)	-0.260
ζ_1 (log(family size))	-24.002
ζ_2 (d. children with age < 6)	-13.903
Measurement error in hours	
σ_ϵ	19.31
Wage equation	
η_1 (const)	1.4
η_2 (log(age/17))	2.75
η_3 (square of log(age/17))	-2.2
Heterogeneity in wages	
τ	0.981
Random preferences	
γ_1	3.0
γ_2	4.0

3.6.1 Monte Carlo, no random preferences, predicted budget constraints

The first estimation method is the estimation of a simplified model which neglects random preferences and ignores random variation of budget constraints for non-participants. In the absence of random preferences the labour supply function (3.14) becomes

$$\bar{h}_n = h(w_n, \mu_n; \alpha, x_n'\omega) \tag{3.63}$$

Optimal labour supply is determined according to scheme (3.59) with $v_n = 0$. The only source of randomness now is measurement error ϵ and the participation rule in this model is:

$$\begin{aligned} h_n^* &= \bar{h}_n + \epsilon_n \\ h &= 0 \quad \text{if} \quad h_n^* \leq 0 \\ &= h_n^* \quad \text{if} \quad h_n^* > 0 \end{aligned} \tag{3.64}$$

[9] This small number of replications is due to the fairly heavy computational burden associated with estimation of the completely specified model

The coherency constraint (3.54) remains the same.

For the non-participants one single budgetconstraint is used, i.e. variation in the budget constraint due to variation in wages is ignored. This means that the wages for non-participants are predicted from the systematic part of the wage equation. For participants, the distributional assumptions (3.55), (3.56) and (3.57) are maintained. To avoid selection bias, the wage equation will be estimated for participants jointly with a selectivity equation of the form

$$y_n^* = \kappa' z_n + e_n \tag{3.65}$$

which can be interpreted as an approximate reduced form of

$$h_n^* = h(\exp(\eta' x_n + u_n), \mu_n; \alpha, x_n' \omega + v_n) + \epsilon \tag{3.66}$$

It is clear that the vector of variables z_n should contain all the variables included in the wage equation, as well as the variables appearing in the labour supply function $h(.)$. The joint wage-participation model is

$$\begin{array}{lcl} \ln w_n & = & \eta' x_n + u_n \\ y_n^* & = & \kappa' z_n + e_n \\ y_n^* & > & 0 \text{ if working } (w_n \text{ observed}) \\ & \leq & 0 \text{ if not } (w_n \text{ unobserved}) \end{array} \tag{3.67}$$

with

$$\begin{pmatrix} \ln \lambda_n \\ e_n \end{pmatrix} \sim N\left(\begin{pmatrix} m_n \\ 0 \end{pmatrix}, \Sigma \right) \tag{3.68}$$

and

$$\Sigma = \begin{pmatrix} \tau^2 & \sigma_{ue} \\ \sigma_{ue} & 1 \end{pmatrix} \tag{3.69}$$

where λ_n and m_n have been defined in (3.55) and (3.57). The variance of e_n has been normalized to one.

Once the model (3.67)-(3.69) has been estimated, (3.63) is estimated with predicted wages for non-participants. The predicted wages come from the systematic part of the wage equation.

Table 3.2 shows the Monte Carlo results for the estimation of this reduced form wage-participation model. The means in column 2 refer to the averages of estimates over 20 replications. Column three shows the standard deviations of the estimates over the replications and column four presents the average of the estimated asymptotic standard errors. Relative errors are given in the final column. They are defined by $|\bar{\theta} - \theta_0|/|\theta_0|$, where $\bar{\theta}$ is the mean in column 2 and θ_0 is the true parameter value. The mean of the parameter estimates $\hat{\eta}_2$ and $\hat{\eta}_3$, which correspond to the age variables in the wage equation are somewhat higher in absolute value than the true parameter values. The same holds for the variance τ. The parameters κ_j of the participation equation all are significant.

The Monte Carlo results for the labour supply parameters are given in table 3.3. The mean estimates of the constant term, α_2, ζ_1 and ζ_2 are all larger in absolute value than their true values, but their sign is estimated correctly. The estimate of α_3, the parameter of the linear wage term in the labour supply function, is close to its true value. However, the mean estimate of α_4, which corresponds to the quadratic wage term in the labour supply function is close to zero as compared with the true value. In fact the mean estimate of α_4 is about five standard deviations below the true value. The standard deviations and the mean SE are fairly similar. The variance σ_ϵ is higher than the true value, which is due to the neglect of random preferences.

3.6.2 Monte Carlo, IV, participants only

The next method we consider is the instrumental variables method. Now the non-convex piecewise linear budget constraint is linearized. Only participants are taken into consideration. We look at the observed value of labour supply, h. We check between which kinkpoints of the budget constraints this value is. Suppose that $H_{j-1} < h < H_j$: Then we are on the j-th segment and the budget constraint is linearized by the linear budget constraint with slope w_j and intercept μ_j that correspond to segment j. The equation we might want to estimate is:

$$h = \alpha_1 + \alpha_2\mu_j + \alpha_3 w_j + \frac{1}{2}\alpha_4 w_j^2 + x'\zeta + \epsilon \tag{3.70}$$

in which $w_j = (1 - \tau_j)w$, with w the gross wage rate and τ_j the marginal tax rate of segment j.

As is well known, there are two reasons why this equation cannot be simply estimated by OLS. The first is the presence of correlation between (μ_j, w_j) and the error term, and the second is the selectivity problem.

There are two causes for correlation between (μ_j, w_j) and the error term ϵ. The first is that the value of h determines which segment of the budget constraint is the appropriate segment, and consequently it determines which pair (μ_j, w_j) is chosen. Secondly, optimal labour supply need not coincide with observed labour supply due to measurement error. As the choice of the segment is determined by observed labour supply, instead of optimal labour supply, the wrong segment may be chosen. As a consequence, (μ_j, w_j) will be subject to measurement error as well, and their measurement errors are correlated with the measurement error of labour supply. The fact that we do not observe individuals at kink points can also be explained by measurement error. Instrumental variables for the intercept μ_j, the slope w_j and its square w_j^2 will have to be used. Obvious candidates are non-labour income μ, the gross wage rate w and its square, and individual characteristics that appear in the wage equation. It has to be assumed that the gross wage rate and ϵ are uncorrelated.

As we restrict ourselves to participants, a selectivity problem arises. We solve this problem by applying the standard Heckman correction: In (3.67) a reduced form wage-participation model has been presented. Suppose that the error term of the gross wage rate, u, is uncorrelated with the error ϵ of the labour supply equation. Next, make the standard assumption (false in this case) that the error term of the participation equation, e, and ϵ are jointly normally distributed. Then the expectation of ϵ, conditional on participation, $y^* > 0$, can be derived:

$$\lambda = \frac{\phi(-\kappa' z)}{1 - \Phi(-\kappa' z)} \tag{3.71}$$

in which $\phi(.)$ is the standard normal density function and $\Phi(.)$ the standard normal distribution function.[10] The estimate $\hat{\kappa}$, obtained from estimating the reduced form participation model can be used as a value for κ. To correct for selectivity, $\hat{\lambda}$ is added to the labour supply equation, where $\hat{\lambda}$ is equal to λ with κ replaced by $\hat{\kappa}$. The final estimation equation becomes:

$$h = \alpha_1 + \alpha_2\mu_j + \alpha_3 w_j + \frac{1}{2}\alpha_4 w_j^2 + x'\zeta + \sigma_{\epsilon e}\hat{\lambda} + \epsilon \tag{3.72}$$

The model has been estimated on the Monte Carlo data, using two sets of instrumental variables. The *extended set* of instrumental variables contains, apart from the constant term, the correction term and the vector of characteristics x that already appears in the labour supply function, non-labour income μ, the gross wage rate w and its square and the age variables that also appear in the wage equation. These age variables are excluded from the *restricted set* of instrumental variables. Tables 3.4 and 3.5 present the Monte Carlo results obtained with the full set and the restricted set of instrumental variables respectively. There is not much difference between the Monte Carlo results with the different sets of instrumental variables. The signs of the estimates are correct, and there is significant evidence for the backward bending labour supply curve. The parameter α_3 of the linear wage term is a bit underestimated, whereas parameter α_4 of the quadratic wage term is overestimated. The standard deviations of α_2, ζ_1 and ζ_2 are sizeable and the estimates are lower than the corresponding true values.

In the estimation with the instrumental variable methods, no coherency constraints are imposed and therefore these constraints may not be satisfied for all individuals. For the estimates obtained with the restricted set of instrumental variables, 6.3% of the individuals does not satisfy the coherency constraint. For the extended set this percentage is 8.3.

Comparing tables 3.4 and 3.5 with table 3.3, it appears that the instrumental variables method performs somewhat worse on average than the estimation of labour supply with neglect of random preferences, as carried out in the previous subsection, except for the fact that the instrumental variables method does manage to reproduce the backward bending labour supply curve.

3.6.3 Monte Carlo, MSS

We now consider the estimation by means of the method of simulated scores. Random preferences, as well as the tax and social security system are properly accounted for. Two variants can be distinguished. In the first variant we use predicted wages for non-working individuals. The predictors are obtained from the reduced form wage-participation model (3.67). Stochastic variation in the budget constraint due to stochastic variation in the wages is ignored here. The second method consists of estimating parameters of the labour supply model and the wage equation jointly.

In table 3.6 the Monte Carlo results of the variant with predicted wages are given. For the optimization of the objective function the downhill simplex method has been used. The results of the Monte Carlo study of the model without random preferences have

[10]Strictly speaking we only assume (3.71), which is weaker than normality of e and ϵ

been used in the construction of an initial starting simplex. The number of drawings to construct the simulator is equal to 10. The matrix of instruments has been constructed on basis of the true parameter values, using $R_Z = 800$ drawings. (Recall that the matrix of instruments has to be calculated only once at the beginning of the optimization procedure.) The consequences of constructing the matrix of instruments with estimated values in the context of a two step estimation procedure have been studied in Bloemen and Kapteyn (1993a). It was found that the means of the estimated values do not change much by employing the two step procedure, but the standard deviations of the estimates are higher than in the case in which the matrix of instruments is calculated based on the true parameter vector.

The present method of estimation clearly outperforms the previous two estimation methods. Although the parameter of the quadratic wage term in the labour supply function is still underestimated, it is closer to its true value as well, and it is significantly different from zero at the 10% level.

Table 3.7 presents the Monte Carlo results of jointly estimating the labour supply model and the wage equation. Again, $R_Z = 800$ drawings were used to contruct the matrix of instruments and $R = 10$ drawings were used to construct the simulators. The estimates of the labour supply parameters are not very different from those obtained with the nonstochastic budget constraint in table 3.6. There is, however, a slight improvement in the estimation of the parameter of the quadratic wage term, which is closer to its true value than for any of the previous methods of estimation. Comparing the parameters of the wage distribution with the parameters obtained with the reduced form wage participation model in table 3.2, we see that there is an improvement in all but one of the parameters estimates.

In conclusion we may say that the Monte Carlo results show that the method of simulated scores with a nonstochastic budget constraint already yields fairly reasonable results, as compared to approximate methods like the instrumental variables method or leaving out random preferences. All of the methods of estimation seem to have problems in properly estimating the parameter of the quadratic wage term in the labour supply function. Leaving out random preferences severely underestimates the parameter of the quadratic wage term, whereas the instrumental variables method overestimates this parameter. In the Monte Carlo study the joint estimation of the labour supply function and the wage equation gives the best results with respect to the quadratic wage effect.

3.6.4 Estimation, no random preferences, predicted budget constraints

The model without random preferences is estimated using the 1985 OSA data, which includes 849 married female individuals of which 331 have a paid job. First the wage-participation model is estimated and the estimates are presented in table 3.8. Apart from the age variables, four education dummies have been included in the wage equation and consequently also in the participation equation. Educl is a dummy variable for the lowest level of education. The highest level of education is taken as the reference category. The four education dummies in the wage equation are negative and significant and they are increasing with the level of education, as they should. The age-earnings profile reaches its maximum at the age of 36. The dummy for the number of children with age below 6 and log family size have a significant negative effect on participation. A higher level of education tends to reduce the probability of non-participation. The probability of participation rises with age until the age of 29 after which it decreases. The covariance σ_{ue} between wages and participation is insignificant.

The wage estimates are used to predict wages for the non-participants, after which the labour supply model without random preferences is estimated.

The parameter estimates are given in table 3.9. Non-labour income has a small but significant (at the 10% level) negative effect on labour supply. The parameter estimate of α_3, which corresponds to the linear wage term in the labour supply equation is positive and significant at the 10% level. The quadratic wage term does not seem to play a significant role in the labour supply function, so the present estimates do not provide evidence for a backward bending labour supply curve. Both the presence of children with age below 6 and an increase in log family size have a significant negative effect on labour supply. The age variables turn out to be insignificant.

3.6.5 Estimation, IV, participants only

Tables 3.10 and 3.11 present the empirical results, obtained with the instrumental variables method. Again two sets of instrumental variables are used. The *restricted set* contains the constant term, the dummy for the presence of children with age below 6, log family size, the correction term and the age variables. The *full set* contains, in addition to the variables that have been included in the restricted set, the education dummies which also appear in the wage equation.

There is not much difference between the IV estimates obtained with the restricted set and the IV estimates obtained with the full set of instrumental variables. The difference with the estimates in table 3.9, obtained by the model without random preferences, are considerable. Non-labour income has a larger impact on labour supply according to the IV estimates. Remarkably the IV estimates provide evidence in favour of a backward bending labour supply curve, as opposed to the estimates in table 3.9 in which the parameter estimate of α_4 was insignificant. This is in line with the Monte Carlo results presented above. Also there IV generated by far the largest estimate (in absolute value) of the quadratic wage effect. According to the IV estimates, the dummy for presence of children with age below 6 has a positive, though insignificant, effect. Log family size still has a significantly negative effect, but its estimated impact on labour supply is much smaller than according to table 3.9. The age variables are insignificant for both types of estimators.

The percentage of individuals that does not satisfy the coherency constraint is 45 for the restricted set and 43 for the extended set of instrumental variables.

3.6.6 Estimation, MSS

Table 3.12 shows the estimation results with MSS and a non-stochastic budget constraint.[11] The standard errors of the individual characteristics in the labour supply function are high relative to the estimates.

Table 3.13 presents the estimates obtained with the method of simulated scores and a stochastic budget constraint. The estimate of α_3, the parameter of the linear wage term in the labour supply function, is larger than the estimates for this parameter that we obtained with the IV and no random preferences methods. Non-labour income has a larger impact as well. The standard errors of the parameters of the wage distribution are rather high. The same holds, to a lesser extent, for the the standard errors of the parameters ζ_i. Comparing the estimates of the wage parameters obtained with MSS with the estimates obtained with the reduced form wage-participation model in table 3.8, it is clear that there are differences. In view of the large standard errors in table 3.13, this does not necessarily mean much. Again, the quadratic wage term in the labour supply function is not significant. The parameter estimate of γ_2 is of large magnitude and estimated imprecisely.

3.6.7 Wage and participation elasticities for different methods

Table 3.14 shows the wage and participation elasticities that are implied by the various methods of estimation. These have been calculated as "aggregate" elasticities in the sense that all wages in the sample have been raised by 5% and then hours and participation have been predicted for every individual in the sample. For the method of simulated scores with a stochastic budget constraint the wages for non-working individuals have been simulated using the estimates of the wage distribution in table 3.13. For the remaining estimation methods the predicted wages based on the estimates in table 3.8 have been used. For the simulation of hours and participation the scheme (3.59) with $v_n = 0$ has been used for the IV method and the method without random preferences, whereas simulated values for v_n have been inserted for the MSS methods. The wage elasticities range from 0.11 for MSS with a non-stochastic budget constraint up to 1.29 for the model without random preferences. The participation elasticities range from 0.064 to 0.99. The elasticities with the IV method have been calculated both including and excluding the individuals that do not satisfy their coherency constraint. The standard errors of the estimated elasticities are sizeable. Consequently, the differences between the elasticities are not significant.

Table 3.15 presents the wage and participation elasticities of the Monte Carlo data. Also, for the Monte Carlo data, the different estimation results produce different elasticities, although the variation is a bit less than for the real data. Note that the ranking of the elasticities by method of estimation coincides with the ranking of the empirical elasticities in table 3.14, except for MSS with a non-stochastic budget constraint, which exhibits much larger elasticities for the Monte Carlo data than for the real data. The

[11] To be sure, throughout we assume that the budget constraint is non-stochastic from the viewpoint of the agent; however from the viewpoint of the econometrician the budget constraint is stochastic since we do not observe all sources of heterogeneity across individuals

standard errors of the elasticities of the Monte Carlo data are smaller than for the real data. For the Monte Carlo data, most of the differences in the elasticities are significant.

Altogether, it is clear that there are considerable differences in the empirical estimates obtained by the various methods of estimation. These differences in the estimates have implications for the wage and participation elasticities, which vary widely across different methods of estimation. The standard errors of the elasticities are sizeable. Apart from the IV method, none of the methods provide evidence in favour of a backward bending labour supply curve. In the Monte Carlo experiment we saw that IV tends to overestimate the quadratic wage term in the labour supply function, whereas the other methods had a tendency to underestimate.

Table 3.2 Monte Carlo: Wage-participation model

θ	true value	mean	SD	mean SE	rel. err.
Participation equation (3.65)					
κ_1 (const)	–	0.825	0.203	0.251	–
κ_2 (log(fam. size))	–	-0.536	0.0552	0.0864	–
κ_3 (d. child. < 6)	–	-0.264	0.0604	0.0639	–
κ_4 (non-labour income)	–	-0.000819	0.0000940	0.0000742	–
κ_5 (log(age/17))	–	3.083	0.699	0.798	–
κ_6 (square of log(age/17))	–	-2.425	0.516	0.544	–
Wage equation (3.67)					
η_1 (const)	1.4	1.097	0.169	0.235	0.22
η_2 (log(age/17))	2.75	3.408	0.564	0.698	0.24
η_3 (square of log(age/17))	-2.2	-2.646	0.440	0.483	0.20
τ	0.981	1.145	0.0591	0.0388	0.17
σ_{ue}	–	1.136	0.0632	0.0406	–

Table 3.3 Monte Carlo: Labour supply model
No random preferences

θ	true value	mean	SD	mean SE	rel. err.
α_1 (const)	14.1	15.768	5.734	9.456	0.12
α_2 (non-labour income)	-0.047	-0.0631	0.00980	0.00648	0.34
α_3 (wage)	10.69	10.704	1.286	2.270	0.0014
α_4 ($0.5 \times$ square of wage)	-0.26	-0.0126	0.0477	0.115	0.95
ζ_1 (log(family size))	-24	-34.044	5.218	3.788	0.42
ζ_2 (d. children with age < 6)	-13.9	-17.594	5.607	3.328	0.27
σ_ϵ	19.3	24.234	0.729	0.856	0.26

Table 3.4 Monte Carlo: Labour supply model
The Instrumental Variables method
Extended set of Instrumental Variables

θ	true value	mean	SD	mean SE	rel. err.
α_1 (const)	14.1	28.657	6.808	7.618	1.03
α_2 (non-labour income)	-0.047	-0.0284	0.0207	0.0209	0.40
α_3 (wage)	10.69	7.391	1.991	2.210	0.31
α_4 (0.5 $\times$ square of wage)	-0.26	-0.355	0.160	0.158	0.36
ζ_1 (log(family size))	-24	-11.236	7.151	5.354	0.53
ζ_2 (d. children with age < 6)	-13.9	-5.868	3.481	3.905	0.59
σ_ϵ	19.3	27.352	2.907	—	0.42

Table 3.5 Monte Carlo: Labour supply model
The Instrumental Variables method
Restricted set of Instrumental Variables

θ	true value	mean	SD	mean SE	rel. err.
α_1 (const)	14.1	26.591	6.934	8.234	0.88
α_2 (non-labour income)	-0.047	-0.0351	0.0244	0.0232	0.25
α_3 (wage)	10.69	8.071	1.965	2.401	0.24
α_4 (0.5 $\times$ square of wage)	-0.26	-0.377	0.165	0.170	0.45
ζ_1 (log(family size))	-24	-12.069	7.826	5.677	0.50
ζ_2 (d. children with age < 6)	-13.9	-6.567	3.940	4.176	0.53
σ_ϵ	19.3	28.623	3.707	—	0.48

Table 3.6 Monte Carlo: Labour supply model
Method of Simulated Scores, R = 10
Non-stochastic budget constraint

θ	true value	mean	SD	SE	rel. err.
α_1 (const)	14.1	13.639	9.573	7.683	0.035
α_2 (non-labour income)	-0.047	-0.0504	0.0116	0.00335	0.075
α_3 (wage)	10.69	11.041	1.182	1.004	0.033
α_4 (0.5 $\times$ square of wage)	-0.26	-0.169	0.101	0.360	0.35
ζ_1 (log(family size))	-24	-25.693	9.071	7.351	0.070
ζ_2 (d. children with age < 6)	-13.9	-14.604	6.032	7.409	0.050
σ_ϵ	19.31	20.851	2.625	6.019	0.080
γ_1	3.0	3.025	1.319	2.239	0.0082
γ_2	4.0	4.111	0.883	2.561	0.11

Table 3.7 Monte Carlo:
Labour supply model and wage distribution
Method of Simulated Scores, R = 10
Stochastic budget constraint

θ	true value	mean	SD	SE	rel. err.
α_1 (const)	14.1	14.467	8.829	4.585	0.023
α_2 (non-labour income)	-0.047	-0.0517	0.0129	0.0138	0.10
α_3 (wage)	10.69	11.075	1.452	0.903	0.036
α_4 (0.5 $\times$ square of wage)	-0.26	-0.176	0.0967	0.210	0.32
ζ_1 (log(family size))	-24	-25.409	8.048	5.831	0.059
ζ_2 (d. children with age < 6)	-13.9	-14.750	4.112	4.107	0.061
σ_ϵ	19.31	23.966	5.475	2.481	0.24
γ_1	3.0	3.222	1.076	0.403	0.074
γ_2	4.0	4.711	2.176	1.061	0.18
The wage distribution					
η_1 (const)	1.4	1.444	0.365	0.197	0.044
η_2 (log(age/17))	2.75	3.395	0.547	0.439	0.23
η_3 (square of log(age/17))	-2.2	-3.149	0.812	0.334	0.43
τ	0.981	1.082	0.103	0.0195	0.10

Table 3.8 Estimates of the wage-participation model

θ	$\hat{\theta}$	SE
κ_1 (const)	2.922	0.531
κ_2 (log(fam. size))	-1.335	0.179
κ_3 (d. child. < 6)	-1.085	0.150
κ_4 (non-labour income)	-0.000340	0.000163
κ_5 (log(age/17))	3.136	1.108
κ_6 (square of log(age/17))	-2.970	0.739
κ_7 (educ1)	-1.819	0.435
κ_8 (educ2)	-1.983	0.435
κ_9 (educ3)	-1.521	0.426
κ_{10} (educ4)	-0.915	0.456
η_1 (const)	2.493	0.164
η_2 (log(age/17))	1.896	0.470
η_3 (square of log(age/17))	-1.277	0.335
η_4 (educ1)	-0.668	0.0941
η_5 (educ2)	-0.562	0.0847
η_6 (educ3)	-0.477	0.0564
η_7 (educ4)	-0.213	0.0609
τ	0.470	0.0134
σ_{ue}	0.0396	0.0743

Table 3.9 Estimates of the labour supply model
No random preferences

θ	$\hat{\theta}$	SE
α_1 (const)	15.049	11.819
α_2 (non-labour income)	-0.00720	0.00369
α_3 (wage)	3.200	1.745
α_4 (0.5 $\times$ square of wage)	-0.0315	0.127
ζ_1 (log(family size))	-30.995	4.563
ζ_2 (d. children with age < 6)	-22.787	3.645
ζ_3 (log(age/17))	2.726	26.15
ζ_4 (square of log(age/17))	-24.578	17.503
σ_ϵ	24.240	1.540

Table 3.10 Estimates of the labour supply model
Instrumental Variables: restricted set

θ	$\hat{\theta}$	SE
α_1 (const)	35.842	4.218
α_2 (non-labour income)	-0.0389	0.0237
α_3 (wage)	1.719	0.422
α_4 (0.5 $\times$ square of wage)	-0.0950	0.0252
ζ_1 (log(family size))	-6.422	2.948
ζ_2 (d. children with age < 6)	0.748	2.487
ζ_3 (log(age/17))	-18.154	11.957
ζ_4 (square of log(age/17))	4.684	8.963
σ_ϵ	9.722	—
Correction term	-3.380	2.543

Table 3.11 Estimates of the labour supply model
Instrumental Variables: Full set

θ	$\hat{\theta}$	SE
α_1 (const)	35.780	4.175
α_2 (non-labour income)	-0.0372	0.0235
α_3 (wage)	1.719	0.417
α_4 (0.5 $\times$ square of wage)	-0.0965	0.0247
ζ_1 (log(family size))	-6.307	2.938
ζ_2 (d. children with age < 6)	0.870	2.478
ζ_3 (log(age/17))	-18.280	11.928
ζ_4 (square of log(age/17))	4.770	8.941
σ_ϵ	9.699	—
Correction term	-3.498	2.535

Table 3.12 Estimates of the labour supply model
Method of Simulated Scores, R = 10
Non-stochastic budget constraint

θ	$\hat{\theta}$	SE
α_1 (const)	37.867	29.733
α_2 (non-labour income)	-0.0194	0.0140
α_3 (wage)	11.522	6.790
α_4 (0.5 $\times$ square of wage)	-0.00462	0.468
ζ_1 (log(family size))	5.902	4.800
ζ_2 (d. children with age $<$ 6)	14.900	16.738
ζ_3 (log(age/17))	85.328	70.310
ζ_4 (square of log(age/17))	-21.219	12.290
σ_ϵ	16.547	4.966
γ_1	4.370	2.187
γ_2	30.871	22.872

Table 3.13 Estimates of the
labour supply model and wage distribution
Method of Simulated Scores, R = 10
Stochastic budget constraint

θ	$\hat{\theta}$	SE
α_1 (const)	31.145	23.677
α_2 (non-labour income)	-0.0480	0.0216
α_3 (wage)	10.285	3.729
α_4 (0.5 $\times$ square of wage)	-0.00243	0.318
ζ_1 (log(family size))	-26.251	19.765
ζ_2 (d. children with age $<$ 6)	-17.912	21.983
ζ_3 (log(age/17))	56.486	40.756
ζ_4 (square of log(age/17))	-36.518	29.024
σ_ϵ	12.429	2.877
γ_1	4.138	2.022
γ_2	23.877	24.309
The wage distribution		
η_1 (const)	0.630	1.101
η_2 (log(age/17))	2.666	1.832
η_3 (square of log(age/17))	-1.325	1.477
η_4 (educ1)	-0.935	0.998
η_5 (educ2)	-0.859	0.980
η_6 (educ3)	-0.719	0.825
η_7 (educ4)	-0.581	0.942
τ	0.653	0.0192

Table 3.14 Wage and participation elasticities

Method of estimation	wage elasticity	SE	participation elasticity	SE
No random preferences	1.29	1.17	0.99	1.24
IV, restricted set	0.23	0.70	0.15	0.74
IV, restricted set (excluding non-coherents)	0.21	0.69	0.16	0.73
IV, full set	0.19	0.70	0.12	0.67
IV, full set (excluding non-coherents)	0.19	0.70	0.13	0.68
MSS, non-stochastic budget constraint	0.11	0.05	0.064	0.04
MSS, stochastic budget constraint	0.39	0.43	0.29	0.21

Table 3.15 Monte Carlo: wage and participation elasticities

Method of estimation	wage elasticity	SE	participation elasticity	SE
No random preferences	1.05	0.11	0.73	0.08
IV, restricted set	0.33	0.21	0.12	0.19
IV, restricted set (excluding non-coherents)	0.36	0.19	0.16	0.19
IV, full set	0.26	0.18	0.09	0.15
IV, full set (excluding non-coherents)	0.30	0.16	0.08	0.15
MSS, non-stochastic budget constraint	0.95	0.20	0.61	0.26
MSS, stochastic budget constraint	0.69	0.11	0.40	0.11

3.7 Conclusions

Both the Monte Carlo results and the estimation results for real data show large variation of outcomes across estimation methods. For the Monte Carlo we know the true model and the results suggest that an incorrect treatment of the stochastic nature of the data may lead to large biases. Estimated wage and participation elasticities may easily be double or half the true elasticity if the wrong estimation method is applied.

For the real data, we do not know the true model, of course, but the huge variation in parameters and implied elasticities is disconcerting. The fact that the ordering of elasticities is by and large the same as for the Monte Carlo data is suggestive of the fact that also here an oversimplification of stochastic structure may be a cause of biased outcomes.

In itself the model considered in this chapter cannot claim to be realistic. After all, it does not have any dynamic elements, no fixed costs of working, etc. The purpose of this chapter has not been to build a fully realistic model of labour market behavior. Rather we have limited ourselves to a somewhat simplified environment in which agents are supposed to behave and then concentrated on a utility consistent specification behavior in that environment. Where our results seem to show the extreme importance of a correct (utility consistent) treatment of the stochastic structure of the model in such a case, we would anticipate even more relevance of such treatment in more complicated environments.

3.A Simulation of the score

In this appendix the technical details of the simulation of the score will be worked out. The simulation of the score can be split up in two parts, i.e. the simulation of the participation probabilities and the simulation of the score of the continuous part of the likelihood function.

First, some notation is introduced. Let μ_{nj} denote the intercept of the j-th segment of the budget constraint, as indicated in figure 3.4, where $j = 1, ..., m$. The index $j = 0$ indicates the segment which introduces the non-convexity in the budget constraint. The slope of the j-th segment is denoted by w_{nj}, $w_{n0} < w_{n1}, w_{nj} > w_{n,j+1}, j = 1, ..., m-1$, and H_{nj} is the kink point between the j-th and $(j+1)$-th segment, $j = 0, ..., m-1$. If $H_{n0} = 0$ we just have the model without social security system. If we allow for variation in the gross wage w_n, the slopes w_{nj} and the kink points H_{nj} will depend on w_n, so formally:

$$
\begin{aligned}
w_{nj} &= w_j(w_n) &(3.A.1)\\
H_{nj} &= H_j(w_n) &(3.A.2)\\
w_j'(w_n) &> 0 &(3.A.3)\\
H_j'(w_n) &< 0 &(3.A.4)
\end{aligned}
$$

As in section 3.5, we let h_n^j denote the optimal amount of labour supply if the budget constraint is linear with slope w_{nj} and intercept μ_{nj}, $j = 1, ..., m$. In our model:

$$h_n^j = h^\bullet(\alpha, w_{nj}, \mu_{nj}, \zeta, x_n) + \alpha_2 v_n \tag{3.A.5}$$

where

$$h^\bullet(\alpha, w, \mu, \zeta, x) = \alpha_1 + \alpha_2\mu + \alpha_3 w + \frac{1}{2}\alpha_4 w^2 + x'\zeta \tag{3.A.6}$$

Expressing α and ζ in terms of the original parameters gives:

$$
\begin{aligned}
\alpha_1 &= \delta + \beta\theta_0 &(3.A.7)\\
\alpha_2 &= \beta &(3.A.8)\\
\alpha_3 &= \gamma + \beta\delta &(3.A.9)\\
\alpha_4 &= \beta\gamma &(3.A.10)\\
\zeta &= \omega\beta &(3.A.11)
\end{aligned}
$$

Because $\beta < 0$ and $\gamma > 0$ we find that $\alpha_2 < 0$ and $\alpha_4 < 0$. Notation will be abbreviated by defining

$$h_{nj}^\bullet = h^\bullet(\alpha, w_{nj}, \mu_{nj}, \zeta, x_n) \tag{3.A.12}$$

The unobserved taste parameter v_n is assumed to be distributed according to a negative gamma-distribution with parameters γ_1 and γ_2, i.i.d. over individuals, by which we mean that $-v_n$ has a gamma distribution. The probability density function of v_n is

$$g(v_n, \gamma_1, \gamma_2) = \frac{1}{\Gamma(\gamma_1)\gamma_2^{\gamma_1}}(-v_n)^{\gamma_1 - 1}\exp\left(\frac{v_n}{\gamma_2}\right), \gamma_1 > 0, \gamma_2 > 0, -\infty < v_n < 0 \tag{3.A.13}$$

As pointed out in section 3.5, the distribution of the measurement errors is assumed to be normal with mean zero and variance σ_ϵ^2:

$$\phi(\epsilon_n, \sigma_\epsilon^2) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\epsilon_n^2\right\}, -\infty < \epsilon_n < \infty \tag{3.A.14}$$

The wage distribution can be derived from assumptions (3.55), (3.56) and (3.57).

$$\begin{aligned} \kappa(w_n, \eta, \tau^2) &= \frac{1}{\sqrt{2\pi}\tau} \frac{\tilde{w}}{\tilde{w} - w_n} \frac{1}{w_n} \exp\left\{-\frac{1}{2\tau^2}\left[\log\left(\frac{w_n}{\tilde{w} - w_n}\right) - m_n\right]^2\right\} && (3.A.15) \\ m_n &= -\log\left(\frac{\tilde{w}}{\exp(\eta' x_n)} - 1\right), 0 < w_n < \tilde{w} && (3.A.16) \end{aligned}$$

It is a straightforward extension to incorporate correlation between wages and measurement errors. However, to restrict the introduction of notation, we abstain from it here.

The likelihood contribution of an individual will be formulated now. We make use of the scheme (3.59) for determining optimal labour supply and the participation rules described in (3.16) and (3.17). First note that the h_n^j in (3.59) all depend on the random preference parameter v_n, so given everything else, v_n determines in which segment of the budget constraint labour supply is optimal. Therefore, we have to determine which set of values of v_n coincides with which segment of the budget constraint. The following sets are defined:

$$\begin{aligned} A_j &= \{v_n | H_{n,j-1} < h_n^j \leq H_{nj}\} && j = 0, ..., m \\ B_0 &= \{v_n | h_n^0 \leq 0\} \\ B_j &= \{v_n | h_n^j > H_{nj} > h_n^{j+1}\} && j = 1, ..., m-1 \\ B_m &= \{v_n | h_n^m \geq T\} \end{aligned}$$

$$H_{n,-1} = 0, H_{nm} = T$$

$$\begin{aligned} Q(h_{n,NC}, h_{nj}) &= \{v_n | U(h_{n,NC}, y_0(h_{n,NC})) < U(h_{nj}, y_j(h_{nj}))\} \\ Q^*(h_{n,NC}, h_{nj}) &= \{v_n | U(h_{n,NC}, y_0(h_{n,NC})) > U(h_{nj}, y_j(h_{nj}))\} \end{aligned}$$

with $h_{n,NC}$ defined in (3.59) and $\quad y_j(h) = w_j h + \mu_j$

$$R_j(h_{n,NC}, h_n^j) = A_j \cap Q^*(h_{n,NC}, h_n^j) \qquad j = 1, ..., m$$

$$
\begin{aligned}
S_j(h_{n,NC}) &= B_j \cap Q^*(h_{n,NC}, H_{nj}) && j = 1, ..., m \\
Z_{01} &= B_0 \cap \left\{ \left(\cup_{j=1}^m R_j(0, h_n^j) \right) \cup \left(\cup_{j=1}^m S_j(0) \right) \right\} \\
Z_{02} &= A_0 \cap \left\{ \left(\cup_{j=1}^m R_j(h_n^0, h_n^j) \right) \cup \left(\cup_{j=1}^m S_j(h_n^0) \right) \right\} \\
Z_{j1} &= B_j \cap Q(h_{n,NC}, H_{nj}) && j = 1, ..., m \\
Z_{j2} &= A_j \cap Q(h_{n,NC}, h_n^j) && j = 1, ..., m \\
Z^* &= \left(\cup_{j=1}^m Z_{j1} \right) \cup \left(\cup_{j=0}^m Z_{j2} \right)
\end{aligned}
\tag{3.A.17}
$$

Z_{01} is the set of v_n for which optimal labour supply is zero, Z_{02} is the set for which it is optimal to be in the first segment of the budget constraint, before the non-convexity kink H_{n0}, Z_{j1} is the set for which optimal labour supply is equal to the j-th kink, $j = 1, ..., m$, Z_{j2} is the set for which it is optimal to be on the j-th segment after the non-convexity kink, $j = 1, ..., m$ and Z^* is the set for which optimal labour supply is positive.

We now determine the probability that observed labour supply is zero, conditional on the value of v_n. According to (3.17) there are two possibilities for observed labour supply to be zero. The first happens when optimal labour supply is zero. Then observed labour supply is equal to zero, irrespective of the value of measurement error. So if v_n is from the set for which optimal labour supply is zero, the probability that observed labour supply is zero, conditional on v_n, is equal to one. The second possibility for observed labour supply to be zero occurs when optimal labour supply is positive but optimal labour supply plus measurement error is negative. Summarizing, the probability that observed labour supply is zero, conditional on v_n becomes:

$$
\begin{aligned}
P(h_n = 0 | v_n, w_n) &= 1 && \text{if} \quad v_n \in Z_{01} \\
&= \Phi\left(-\tfrac{H_{nj}}{\sigma_\epsilon}\right) && \text{if} \quad v_n \in Z_{j1}, j = 1, ..., m \\
&= \Phi\left(-\tfrac{h_n^j}{\sigma_\epsilon}\right) && \text{if} \quad v_n \in Z_{j2}, j = 0, ..., m
\end{aligned}
\tag{3.A.18}
$$

in which $\Phi(.)$ is the standard normal distribution function.

The contribution of positive values of labour supply, conditional on v_n, is restricted to $v_n \in Z^*$ for which optimal labour supply is positive.

$$
\begin{aligned}
\chi(h_n | v_n, w_n) &= \phi(h_n - H_{nj}, \sigma_\epsilon^2) && \text{if} \quad v_n \in Z_{j1}, j = 1, ..., m \\
\chi(h_n | v_n, w_n) &= \phi(h_n - h_n^j, \sigma_\epsilon^2) && \text{if} \quad v_n \in Z_{j2}, j = 0, ..., m
\end{aligned}
\tag{3.A.19}
$$

Having determined the density of observed labour supply, conditional on v_n, the unconditional contribution can be obtained by integrating over v_n.

$$
\begin{aligned}
P(h_n = 0 | w_n) &= \textstyle\int_{Z^* \cup Z_{01}} P(h_n = 0 | v, w_n) g(v, \gamma_1, \gamma_2) dv && \text{if } h_n = 0 \\
l(h_n | w_n) &= \textstyle\int_{Z^*} \chi(h_n | v, w_n) g(v, \gamma_1, \gamma_2) dv && \text{if } h_n > 0
\end{aligned}
\tag{3.A.20}
$$

or, making use of (3.A.18) and (3.A.19)

$$P(h_n = 0|w_n) =$$

$$\int_{Z_{01}} g(v,\gamma_1,\gamma_2)dv + \sum_{j=1}^{m} \int_{Z_{j1}} \Phi\left(-\frac{H_{nj}}{\sigma_\epsilon}\right) g(v,\gamma_1,\gamma_2)dv + \sum_{j=0}^{m} \int_{Z_{j2}} \Phi\left(-\frac{h_n^j}{\sigma_\epsilon}\right) g(v,\gamma_1,\gamma_2)dv \tag{3.A.21}$$

$$\text{if } h_n = 0$$

$$l(h_n|w_n) = \sum_{j=1}^{m} \int_{Z_{j1}} \phi(h_n - H_{nj}, \sigma_\epsilon^2) g(v,\gamma_1,\gamma_2)dv + \sum_{j=0}^{m} \int_{Z_{j2}} \phi(h_n - h_n^j, \sigma_\epsilon^2) g(v,\gamma_1,\gamma_2)dv \tag{3.A.22}$$

$$\text{if } h_n > 0$$

For an individual whose labour supply is zero, wages are not observed and they are integrated out. The final response probability becomes

$$\int_0^{\tilde{w}} P(h_n = 0|w)\kappa(w,\eta,\tau^2)dw \tag{3.A.23}$$

The problem with the above defined sets is that the bounds of these sets are not known explicitly. The advantage of the frequency simulator in the context of an integral with bounds that are known implicitly only, is that it is possible to draw random numbers and then check in which region the simulated value of labour supply is.[12]

We now describe the construction of the frequency simulator. The first thing we need is drawings from the distributions of measurement errors, wages and random preferences. As measurement errors are normally distributed, a series of R random numbers can be drawn from the standard normal distribution which will be kept constant during the minimization process. These basic drawings can be transformed to drawings from the distribution of ϵ_n through multiplying by σ_ϵ. Any change in the drawings of ϵ_n is caused by a change in σ_ϵ.

To draw a series of gross wages we also start by drawing a series of R standard normal random variables $l_{nr}^*, r = 1, ..., R$, which are the constant basic drawings. These basic drawings can be transformed to drawings of the wage rate:

$$w_{nr}^* = \frac{\tilde{w}\exp(m_n + \tau l_{nr}^*)}{1 + \exp(m_n + \tau l_{nr}^*)}, r = 1, ..., R \tag{3.A.24}$$

The transformation is continuous in the parameters and therefore, keeping the basic drawings constant, a change in the drawings w_{nr}^* can only be caused by a change in the parameters.

[12] Note that it is possible to simulate (3.A.23) by drawing wages and random preferences from their respective distributions, without drawing measurement error, then checking the region (Z_{j1} or Z_{j2}) and setting the contribution to the simulator equal to the conditional probability corresponding to this region, evaluated in the drawings. This results in a piecewise continuous simulator, which is a combination of a frequency simulator and a smooth simulator. The possibility to construct this type of simulator however, depends strongly on the model structure imposed in (3.15)-(3.17), which would change if dynamic elements, fixed cost of working, separation of the labour supply decision and participation decision, etc. were introduced. Therefore, the general applicable frequency simulator is employed here.

The generation of random numbers from the negative gamma distribution is not as straightforward as the generation of random numbers from a normal distribution. The method commonly used for the generation of gamma random numbers is the acceptance-rejection method. Although this method is very useful for generating gamma random numbers if the parameters remain constant, the use of this method in the context of a minimization problem with changing parameters is not appropriate. A change in the parameters can cause discrete jumps in the drawings. The alternative would be to generate random numbers by means of the inversion method, see e.g. Devroye (1986). A major drawback of this method is that for every draw the negative gamma distribution function has to be inverted using numerical methods. Experiments with the inversion method have shown that the application of this method in the context of an estimation problem leads to an infeasibly high computational burden, even in rather simple problems. Therefore, the inversion method applied in estimation by simulation procedures is only useful either if the functional form of the inverse of the distribution function is known, or if a good approximation for the inverse of the distribution function is available. A third possibility is to use importance sampling. In that procedure the random numbers are drawn from a different distribution with favourable characteristics and it is corrected for drawing from a different distribution by the use of a weight function. This is the procedure which we use here. We draw random numbers from the exponential distribution with parameter ρ:

$$\Lambda(\rho, v_n) = \rho \exp\{\rho v_n\}, -\infty < v_n < 0, \rho > 0 \tag{3.A.25}$$

Because this is not the "true" (assumed) distribution, the frequency simulator has to be weighted, like in importance sampling. The weight function $k(v_n, \gamma_1, \gamma_2, \rho)$ is the ratio of the negative gamma density function and the negative exponential density function.

$$k(v_n, \gamma_1, \gamma_2, \rho) = \frac{g(v_n, \gamma_1, \gamma_2)}{\Lambda(\rho, v_n)} = \frac{1}{\Gamma(\gamma_1)\gamma_2^{\gamma_1}\rho}(-v_n)^{\gamma_1 - 1} \exp\left\{\left(\frac{1}{\gamma_2} - \rho\right) v_n\right\} \tag{3.A.26}$$

The fact that we draw from the exponential distribution instead of the gamma distribution increases the variance of the estimator. In the first place we have to choose the parameter ρ in such a way that the variance will be finite and second, the choice of ρ has to make the addition to the variance as small as possible. In the implementation a random number v from the negative exponential density in (3.A.25) is inserted in (3.A.26), so in calculating the mean and the variance of the weight function we do this with respect to the negative exponential density. By construction, the mean of the weight function is always equal to one. Note that if it is drawn from the true density the weight function is identically equal to one and as a consequence the variance of the weight function is equal to zero. Therefore, the larger the deviation of the shape of the approximate density function from the true density function is, the larger the variance will be, see e.g. Kloek and Van Dijk (1978). The expression for the variance is given by:

$$\begin{gathered} E[k(v,\gamma_1,\gamma_2,\rho)]^2 - 1 = \\ \int_{-\infty}^{0} \left[\frac{g(v,\gamma_1,\gamma_2)}{\Lambda(\rho,v)}\right]^2 \Lambda(\rho,v)dv - 1 = \\ \int_{-\infty}^{0} k(v,\gamma_1,\gamma_2,\rho)g(v,\gamma_1,\gamma_2)dv - 1 = \\ \frac{\Gamma(2\gamma_1-1)\left(\frac{\gamma_2}{2-\rho\gamma_2}\right)^{2\gamma_1-1}}{[\Gamma(\gamma_1)]^2\gamma_2^{2\gamma_1}\rho} \end{gathered} \tag{3.A.27}$$

in which

$$\gamma_1 > \frac{1}{2} \tag{3.A.28}$$

$$\rho < \frac{2}{\gamma_2} \tag{3.A.29}$$

This is the difference of the mean of the weight function with respect to the true density function $g(v,\gamma_1,\gamma_2)$ and the mean of the weight function with respect to $\Lambda(v,\rho)$ which is equal to one. (3.A.29) is the necessary condition for the variance to be finite. The smallest variance can be obtained by choosing ρ in such a way that the variance of the weight function is minimized. Solving the first order conditions and checking the second order conditions, it can be found that the variance is minimal for

$$\rho = \frac{1}{\gamma_1\gamma_2} \tag{3.A.30}$$

Note, that condition (3.A.29) is satisfied if condition (3.A.28) is satisfied.

Summarizing, the drawing procedure for v_n is as follows. Draw a series of R random numbers $\tilde{v}^*_{nr}$ from the exponential distribution with parameter ρ. These are our basic drawings. Transform the basic drawings to drawings v^*_{nr} from an exponential distribution with parameter $\frac{1}{\gamma_1\gamma_2}$ by multiplying the basic drawings by $\gamma_1\gamma_2$:

$$v^*_{nr} = \gamma_1\gamma_2\tilde{v}^*_{nr} \tag{3.A.31}$$

Note that this is a continuous transformation in γ_1 and γ_2. These are the final drawings which will be used in the simulation of the labour supply.

Having described the way of generating the required random numbers, we now turn to the simulation of the probability. Using the drawings $(\epsilon^*_{nr}, w^*_{nr}, v^*_{nr})$ the optimal labour supply $\bar{h}_{nr}$ and the observed labour supply h_{nr} can be simulated according to scheme (3.59) and the participation rules (3.16) and (3.17). Then the participation probability can be simulated by a frequency simulator like in (3.60)-(3.62) where (3.60) has to be weighted. The frequency simulator becomes:

$$f_{nr} = k(v^*_{nr},\gamma_1,\gamma_2,\rho) \quad \text{if } h_{nr} > 0 \tag{3.A.32}$$

$$f_{nr} = 0 \quad \text{otherwise} \tag{3.A.33}$$

$$F_{nR} = \frac{1}{R}\sum_{r=1}^{R} f_{nr} \tag{3.A.34}$$

By construction, this is an unbiased simulator for the participation probability.

The estimation method also requires a simulator of the derivatives of the probability with respect to the parameters. Let $F_{nR}(\theta)$ denote the frequency simulator in parameter vector θ. Then the derivative with respect to the k-th component of θ is simulated by a difference interval of frequency simulators:

$$\bar{m}_{nk}(\theta, \epsilon_R^*, v_R^*, w_R^*) = \frac{F_{nR}(\theta + \delta e_k) - F_{nR}(\theta)}{\delta} \tag{3.A.35}$$

where e_k is the k-th unit vector. Because $F_{nR}(\theta + \delta e_k)$ is an unbiased simulator of the participation probability in $\theta + \delta e_k$ and $F_{nR}(\theta)$ is an unbiased simulator of the participation probability in θ, (3.A.35) is an unbiased simulator of the difference interval of the participation probability. Because $F_{nR}(\theta)$ is discontinuous in the parameter vector θ we have to choose δ large enough to ensure that the sum of the difference interval over all individuals and all drawings in (3.30) is not equal to zero. The larger the number of drawings R is, the smaller the value of δ can be. To construct the optimal matrix of instruments, which only has to be calculated once at the beginning of the optimization procedure, a large number of drawings can be used. In our empirical applications we used 800 drawings.

We now turn to the simulation of the continuous part of the score vector. First, we abstract from the problems that arise because the bounds are unknown, and from the problem that we cannot draw directly from the gamma distribution. Note, that it is possible to simulate the integral appearing in the log-likelihood function unbiasedly. Draw random numbers v_{nr}^* from the density $g(v, \gamma_1, \gamma_2)$ restricted to the region Z^* for which optimal labour supply is positive, defined in (3.A.17).

$$v_{nr}^* \sim \frac{g(v, \gamma_1, \gamma_2)}{P(v \in Z^*)}, v_{nr}^* \in Z^* \tag{3.A.36}$$

An unbiased simulator for $l(h_n|w_n)$ is

$$\tilde{l}(h_n|w_n) = P(v \in Z^*)\frac{1}{R}\sum_{r=1}^{R} \chi(h_n|v_{nr}^*, w_n) \tag{3.A.37}$$

or, writing this out:

$$\tilde{l}(h_n|w_n) =$$

$$P(v \in Z^*)\frac{1}{R}\sum_{r=1}^{R}\left\{\sum_{j=0}^{m} I(v_{nr}^* \in Z_{j2})\phi(h_n - h_n^j, \sigma_\epsilon^2) + \sum_{j=1}^{m} I(v_{nr}^* \in Z_{j1})\phi(h_n - H_{nj}, \sigma_\epsilon^2)\right\} \tag{3.A.38}$$

in which h_n^j is computed on the basis of v_{nr}^* and I(.) the indicator function. A simulator for the score contribution then could be obtained by simulating numerator and denominator in

$$\frac{\partial \ln l(h_n|w_n)}{\partial \vartheta} \tag{3.A.39}$$

separately. This introduces a bias in the simulation of the score in the sense that the expectation evaluated in the true parameter vector will not be equal to zero. However, this simulator for the continuous part of the score contribution is piecewise continuous and therefore it does not have the unfavourable characteristics of a probability frequency simulator in the context of simulated maximum likelihood, see e.g. Lerman and Manski (1981).

An additional complication arises from the fact that the bounds of the region Z^* are unknown. Hence we have to draw from a different region $\tilde{Z}^*$ which contains the original region, i.e. $Z^* \subset \tilde{Z}^*$, and which approximates the original region as close as possible. Consider the region

$$\begin{array}{rcr} \tilde{Z}^* & = & \{v \mid -\infty < v < q(\alpha, \zeta, w_n, \mu_n)\} \\ q(\alpha, \zeta, w_n, \mu_n) & = & -h^\bullet_{n1}/\alpha_2 \text{ if } -h^\bullet_{n1}/\alpha_2 < 0 \\ & = & 0 \text{ if } -h^\bullet_{n1}/\alpha_2 > 0 \end{array} \tag{3.A.40}$$

$-h^\bullet_{n1}/\alpha_2$ is the value of v for which h^1_n is equal to zero. The region Z^* of positive optimal labour supply is contained in this region. The simulation procedure now becomes as follows: Draw a random number v^*_{nr} from the negative exponential distribution with parameter ρ:

$$v^*_{nr} \sim \rho \exp\{\rho v^*_{nr}\}, -\infty < v^*_{nr} < 0 \tag{3.A.41}$$

Define ν^*_{nr} as

$$\nu^*_{nr} = v^*_{nr} + q(\alpha, \zeta, w_n, \mu_n) \tag{3.A.42}$$

As a consquence, ν^*_{nr} is a truncated negative exponential variable:

$$\nu^*_{nr} \sim \frac{\rho \exp\{\rho \nu^*_{nr}\}}{P(\nu \in \tilde{Z}^*)}, \nu^*_{nr} \in \tilde{Z}^* \tag{3.A.43}$$

Now we construct a simulator that is a combination between a smooth simulator and a frequency simulator. The simulator $\tilde{l}(h_n|w_n)$ becomes

$$\frac{P(\nu \in \tilde{Z}^*)}{R} \sum_{r=1}^{R} I(\nu^*_{nr} \in Z^*) \chi(h_n|\nu^*_{nr}, w_n) k(\nu^*_{nr}, \gamma_1, \gamma_2, \rho) \tag{3.A.44}$$

Until now, we have considered the integration of the integrals appearing in the numerator and denominator of the score contribution separately. However, it is possible to construct a simulator on the basis of the vector of scores which has expectation zero in the true parameter vector. The drawback of this simulator is that we have to draw the random preference variables from their conditional density, i.e. conditional on the observed value of labour supply. The density function conditional on labour supply contains the same integral which we want to avoid to evaluate. Hence, we will consider an approximation. First consider the score of the log-likelihood contribution with respect to parameters that appear in the integrand only. Note that the derivatives of $g(v, \gamma_1, \gamma_2)$ with respect to its parameters can always be written as a multiple of $g(v, \gamma_1, \gamma_2)$ and therefore they can be treated in the same way as the following case.

The derivative of $l(h_n|w_n)$ with respect to σ^2_ϵ is

$$\frac{\partial l(h_n|w_n)}{\partial \sigma^2_\epsilon} = \int_{Z^*} \frac{\partial \chi(h_n|v, w_n)}{\partial \sigma^2_\epsilon} g(v, \gamma_1, \gamma_2) dv \tag{3.A.45}$$

and

$$\frac{\partial \ln l(h_n|w_n)}{\partial \sigma_\epsilon^2} = \frac{\partial l(h_n|w_n)}{\partial \sigma_\epsilon^2} / l(h_n|w_n) \tag{3.A.46}$$

For expository purposes, we ignore again for the moment the problem of the unknown bounds of Z^*. Now suppose that v_{nr}^* can be drawn from

$$\frac{g(v, \gamma_1, \gamma_2)\chi(h_n|v, w_n)}{l(h_n|w_n)}, v \in Z^* \tag{3.A.47}$$

Then the score contribution can be simulated by

$$\frac{\widetilde{\partial \ln} l(h_n|w_n)}{\partial \sigma_\epsilon^2} = P(v \in Z^*) \frac{1}{R} \sum_{r=1}^{R} \frac{\partial \chi(h_n|v_{nr}^*, w_n)}{\partial \sigma_\epsilon^2} / \chi(h_n|v_{nr}^*, w_n) \tag{3.A.48}$$

Taking expectations with respect to the draws v_{nr}^*, this yields the original score component.

For parameters which appear in the bounds of the integral we also have to differentiate the bounds. Suppose that the bound of the region Z^* is given by b. Optimal labour supply, evaluated in the bound is zero. Taking the derivatives of the bound yields

$$\pm \frac{\partial b}{\partial \vartheta} \phi(h_n, \sigma_\epsilon^2) g(b, \gamma_1, \gamma_1) / l(h_n|w_n) \tag{3.A.49}$$

where the sign is positive or negative, depending on whether it is an upperbound or a lowerbound. Taking expectations with respect to positive values of observed labour supply, i.e. with respect to the density $l(h_n|w_n)/P(h_n > 0)$ yields:

$$\pm \frac{1}{2} \frac{\partial b}{\partial \vartheta} g(b, \gamma_1, \gamma_1) / P(h_n > 0) = \frac{1}{2} \frac{\partial \int_{Z^*} g(v, \gamma_1, \gamma_2) dv}{\partial \vartheta} / P(h_n > 0) \tag{3.A.50}$$

which is one half times the derivative of the probability that optimal labour supply exceeds zero, divided by the probability that observed labour supply exceeds zero. Now the same trick can be applied as in Bloemen and Kapteyn (1993a), which means that the original derivative of the bound is replaced by one half times the probability that optimal labour supply is greater than zero for every individual in the sample, both working and non-working. The derivative of the probability that optimal labour supply is zero can be simulated in the same way as in (3.A.35). Summarizing, for parameters which appear in the bound as well as in the integrand we can use the same kind of simulator as in (3.A.48) and in addition to that, we have to adjust the score contribution with a simulator of the derivative of the probability that optimal labour supply is positive for both working and non-working individuals. The resulting score simulator has expectation zero in the true parameter vector. For the problem of the unobserved bounds of Z^*, the same method can be used as in (3.A.44).

The practical appliability of this score method is restricted by the fact that random draws from the conditional density are required. The method, however, still has the advantage that the discontinuities are averaged out as the simulated score consists of linear contributions. Therefore, it is useful to draw random number from an approximate density. The most straightforward way to draw random numbers in this context is to draw them from their marginal distribution. In that case the score contribution will not be unbiased anymore, i.e. the appropriate moment conditions are used in combination with the wrong draws. This point can be made more clear if, for the moment, we ignore the tax system. Suppose that optimal labour supply is given by the function $h_n(v_n)$ for individual n. The participation scheme (3.16) and (3.17) for observed labour supply is

$$\begin{aligned} h_n &= h_n(v_n) + \epsilon \quad \text{if } h_n(v_n) > 0 \text{ and } h_n(v_n) + \epsilon > 0 \\ &= 0 \quad \text{if } h_n(v_n) = 0 \text{ or } h_n(v_n) + \epsilon \leq 0 \end{aligned} \tag{3.A.51}$$

and

$$\epsilon_n \sim \phi(\epsilon_n, \sigma_\epsilon^2) \tag{3.A.52}$$

as before. Then for positive h_n we have

$$h_n \sim \int_{Z^*} \phi(h_n - h_n(v), \sigma_\epsilon^2) g(v) dv / P(h_n > 0) \tag{3.A.53}$$

in which Z^* is, as before, the region for which optimal labour supply is positive. Now draw v_{nr}^* from its marginal density $g(.)$, whereas h_n can be considered as a draw from (3.A.53). Then the implied density for $\epsilon_{nr}^* := h_n - h_n(v_{nr}^*)$ is

$$\int_{z^*} \int_{z^*} \phi(\epsilon_{nr}^* + h_n(\tilde{v}) - h_n(v), \sigma_\epsilon^2) g(v) g(\tilde{v}) dv d\tilde{v} / [P(\tilde{v} \in Z^*) P(h_n > 0)] \tag{3.A.54}$$

which is not equal to $\phi(\epsilon_{nr}^*, \sigma_\epsilon^2)/P(h_n > 0)$. Rather, it is a weighted average of $\phi(\epsilon_n, \sigma_\epsilon^2)/P(h_n > 0)$. Note, that the implicit draws ϵ_{nr}^* are always in the right region, i.e. $h_n(v_{nr}^*) + \epsilon_{nr}^* > 0$. It can be shown that the parameter estimate for the variance of measurement error is biased upwards.

The Monte Carlo results in section 3.6 are obtained with method (3.A.48) using draws form the marginal density of v. From the results it can be seen that the estimate of the variance of measurement error is indeed biased upwards. However, the estimates of the utility parameters do rather well.

Chapter 4

A model of labour supply with job offer restrictions

4.1 Introduction

In this chapter a model of labour supply is set up in which individuals are restricted in their choice by the job offers that are offered to them by the employers. In the mid-eighties it was recognized that the standard neo-classical labour supply model (see e.g. Heckman (1974), Hausman (1980)), in which wages are fixed and the optimal number of working hours can always be chosen, was not in accordance with reality. Dickens and Lundberg (1985) introduced hours restrictions into the labour supply model assuming that hours arrive from a discrete offer distribution. Tummers and Woittiez (1991) extended the model by making the wage rate dependent on hours. Van Soest, Woittiez and Kapteyn (1990) compared the standard model with the model with hours restrictions using a Dutch data set on labour supply, taking into account the tax system, and their estimation results are relevant empirical evidence in favour of the hours restrictions model. In Blundell, Ham and Meghir (1987) a different approach to restrictions on the labour market is followed. Here the emphasis is on the modelling of involuntary unemployment.

Although wages can vary with the number of hours offered in these models, for example because individuals base their decisions on net wages, the possibility that wages also vary from job offer to job offer because of an independent wage offer effect is not taken into consideration. We will assume that, like in job search theory (see Mortensen (1986) for an overview), wages arrive from a wage offer distribution.

We thus make the assumption that a job offer consists of two characteristics. It has a wage component and an hours component. An individual will choose the job yielding the highest utility level. If all jobs offered generate utility levels less than the utility of not working, the individual is observed to be non-working. The possibility is left open that an individual will receive no job offer at all, so involuntary unemployment may arise. Only wage-hours offers that are accepted will be observed. The distribution of the observed wages will depend on the structure of the model and therefore it is not possible to use a twostep procedure to estimate the labour supply parameters, estimating the parameters of the wage ditribution separately. All parameters of the model have to be estimated simultaneously.

As opposed to previous studies, like Tummers and Woittiez (1991) and Van Soest, Woittiez and Kapteyn (1990), the distribution of the number of job offers is made dependent on individual characteristics. It is tested for the significance of the dependence of the average number of job offers on individual characteristics. Particular attention is paid to the effects on the parameter estimates of the utility function, as the introduction of an individual specific number of job offers turns out to have serious consequences for these estimates.

The model is estimated both with and without hours dependence of the wage equation, which enables us to test for the significance of hours in the wage equation. The estimates of the specification with hours in the wage equation provide a different explanation for observing low frequencies of weekly working hours above the full time level than specifications without hours dependent wages.

Simulation experiments are performed to compare the empirical frequency distribution of working hours with the frequency distribution that is generated by the model. Moreover, it is tested formally whether the simulated hours distribution and the empirical hours distribution are equal.

In section 2 we present the model and formulate the likelihood function. In section 3 the maximum likelihood estimates of various model specifications are presented. Simulation experiments and formal testing procedures are performed. Finally, in section 4, some concluding remarks are made.

4.2 The model

The model assumes that the individual cannot supply the optimal number of working hours which results from maximizing utility subject to a budget constraint, like in the standard labour supply model, in which typically no involuntary unemployment can occur, see e.g. Heckman (1974) and Hausman (1980). Instead, it is assumed that the individual, at a given point in time, receives a random amount of job offers from the employers. A job offer is characterized by a wage rate and a weekly number of working hours. The job offers are compared with each other by their utility level. The individual selects the job with the highest utility level, which is compared with the utility level of not working, after which it is decided whether or not the job is accepted. Note that the model is fully static in the sense that all job offers arrive at a given point in time and the job acceptance decision is made immediately, without taking into account the possibility of future job offers. Therefore, the model can be estimated with cross section data. In chapter 5 of this thesis a sequential job search model is presented in which a job not only has a wage component, as is usual in the standard job search model, see e.g. Mortensen (1986), but has an hours component as well, as in the (static) Dickens and Lundberg (1985) model. In this sequential search model, apart from data on labour supply and the wage rate, use is made of duration data in the estimation of the model, so the availability of a panel survey is a requirement. In making the job acceptance decision the expectations with respect to possible future job offers are taken into acount as well.

The present model provides a link between the static Dickens and Lundberg (1985) model, in which only the number of working hours varies from job offer to job offer, and the standard job search model in which the wage rate is the only job component.

The number of job offers is assumed to be Poisson distributed with parameter λ. An advantage of this specification, as opposed to the binomial specification of Tummers and Woittiez (1991) and Van Soest et al. (1990), is that it can easily be made dependent on individual characteristics. In these previous studies, the effect of individual characteristics, like age and the level of education, on the number of job offers, have typically been ignored. Moreover, in the binomial distribution a fixed maximum number of job offers has to be chosen in advance.

A job offer is modelled as a simultaneous draw of a wage rate w and a weekly number of working hours h from a joint wage hours offer distribution $f(w,h)$.

As in the Dickens and Lundberg (1985) approach we assume that there is a discrete hours offer distribution defined over m fixed numbers of positive hours $h_l, l = 1,...,m$. That is, hours are grouped into m categories, where the probability of drawing from category l is given by p_l. The advantage of this approach is that no heavy restrictions, like single peakedness or symmetry, are placed on the shape of the distribution. The price which has to be paid for this flexibility of the shape of the hours distribution is that labour supply can only take a discrete number of values.

$$P(h = h_l) = p_l, l = 1,...,m \tag{4.1}$$

In the estimation the probabilities can be parametrized as

$$p_l = \frac{\mu_l}{\sum_{k=1}^{m} \mu_k}, \mu_1 = 1 \tag{4.2}$$

in which μ_1 has been normalized to 1 and the remaining μ_l are non-negative.

The wage rate, conditional on $h = h_l$, is assumed to be lognormally distributed. Tummers and Woittiez (1991) also estimated the wage distribution jointly with the labour supply probabilities, but they use a normal wage distribution, thereby not restricting the range of possible wages to positive values. The wage specification becomes

$$\ln w = x_l'\eta + v \tag{4.3}$$

$$v \sim N(0, \sigma_v^2) \tag{4.4}$$

or equivalently

$$w = \exp(x_l'\eta)u \text{ with } u = \exp(v) \tag{4.5}$$

in which the subindex l indicates possible dependence on h_l. The joint job offer density function becomes

$$f(w, h_l) = \frac{1}{\sqrt{2\pi}} \frac{1}{w} \exp\left\{-\frac{1}{2\sigma_v^2}[\ln w - \eta' x_l]^2\right\} p_l, 0 < w < \infty, l = 1,...,m \tag{4.6}$$

For ease of exposition it is assumed for the moment that the budget constraint is linear, ignoring the tax system.

$$y = wh + \mu \tag{4.7}$$

where μ is non-labour income.

The utility function is defined over labour supply h and income y and is indicated by $u(h, y)$. The specification of the utility function is the Hausman (1980) specification. Maximizing this utility function subject to a linear budget constraint yields a labour supply function which is linear in both non-labour income and the wage rate.

$$u(h, y) = -\ln(\gamma - \beta h) - \frac{\beta(h - X\delta - e - \beta y)}{\gamma - \beta h} \tag{4.8}$$

where

- β, γ, δ are parameters, $\beta < 0, \gamma > 0$
- y is disposable income
- h is the number of working hours
- e is an unobserved random taste variable, $e \sim N(0, \sigma_e^2)$
- X is a vector of individual characteristics

At a given point in time, an individual receives n job offers each of them consisting of a wage $w, 0 < w < \infty$ and a number of working hours $h \in \{h_1, ..., h_m\}$. Furthermore, an individual can always choose not to work. The alternative which yields the highest level of utility will be chosen. An individual will be observed to be non-working if the utility level of not working exceeds the utility level of all of the n job offers. The number of job offers is assumed to be a Poisson distributed random variable.

$$p(n) = \frac{\exp(-\lambda)\lambda^n}{n!}, n = 0, 1, ..., \infty \tag{4.9}$$

Note, that it is possible that no jobs are offered at all so that individuals can be involuntarily unemployed.

In order to write down the likelihood function the likelihood contribution of non-working and working individuals will be determined separately. Suppose that for a working individual we observe the wage-hours pair (w_*, h_{l_*}), where $l_* \in \{1, ..., m\}$. The fact that (w_*, h_{l_*}) is observed means that all other job offers, if there are any, are from the set of wage-hours packages which yield at most the same utility level as the observed pair. This set has to be determined. For every level of hours $h_l, l = 1, ..., m$, the set $A_l(e)$ of wages can be determined which includes all wage levels w for which $u(h_l, wh_l + \mu) \leq u(h_{l_*}, w_* h_{l_*} + \mu)$ for a given value of e.

$$A_l(e) := \{w | u(h_l, wh_l + \mu) \leq u(h_{l_*}, w_* h_{l_*} + \mu) | e\} \tag{4.10}$$

The probability $P(w_*, h_{l_*}|e)$ of drawing an arbitrary job offer which yields at most the same utility level as the observed job (w_*, h_{l_*}) equals the probability of drawing a job offer from any of the sets $A_l(e)$, i.e.

$$P(w_*, h_{l_*}|e) := P(u(h, wh + \mu) \leq u(h_{l_*}, w_* h_{l_*} + \mu)|e) = \sum_{l=1}^{m} p_l P(w \in A_l(e)|e) \tag{4.11}$$

Using the distributional assumptions in (4.3) and (4.4) yields

$$P(w \in A_l(e)|e) = \Phi\left(\frac{\ln g_l(e)-\eta' x_l}{\sigma_v}\right) \quad \text{if } g_l(e) > 0 \tag{4.12}$$
$$= 0 \quad \text{if } g_l(e) \leq 0 \tag{4.13}$$

with $\Phi(.)$ the standard normal distribution function and

$$g_l(e) = \frac{(\gamma - \beta h_l)\ln\left(\frac{\gamma-\beta h_l}{\gamma-\beta h_{l_*}}\right)}{\beta^2 h_l} - \frac{(h_{l_*} - h_l)(\gamma - \beta X\delta - \beta e - \beta^2\mu)}{\beta h_l(\gamma - \beta h_{l_*})} + \frac{\gamma - \beta h_l}{\gamma - \beta h_{l_*}}\frac{h_{l_*}}{h_l}w_* \tag{4.14}$$

Note that if $l = l^*$ the first two terms of (4.14) are equal to zero, whereas the last term becomes w^*.

Now assume that there are n job offers $(w_{(j)}, h_{(j)}), j = 1, ..., n$. Only the job with the highest utility level, (w_*, h_{l_*}) is observed if its utility level is higher than $u_0 = u(0, \mu)$, the utility level of not working which we will call the reservation utility level. So

$$\begin{array}{llrcl}
(w_*, h_{l_*}) = (w_{(1)}, h_{(1)}) & \text{if} & u(h_{(1)}, w_{(1)}h_{(1)} + \mu) & > & u(h_{(j)}, w_{(j)}h_{(j)} + \mu) \\
 & & & & j = 2, ..., n \\
 & \text{and} & u(h_{(1)}, w_{(1)}h_{(1)} + \mu) & > & u_0 \\
 & & & & \\
(w_*, h_{l_*}) = (w_{(2)}, h_{(2)}) & \text{if} & u(h_{(2)}, w_{(2)}h_{(2)} + \mu) & > & u(h_{(j)}, w_{(j)}h_{(j)} + \mu) \\
 & & & & j = 1, ..., n, j \neq 2 \\
 & \text{and} & u(h_{(2)}, w_{(2)}h_{(2)} + \mu) & > & u_0 \\
 & \vdots & & & \\
(w_*, h_{l_*}) = (w_{(n)}, h_{(n)}) & \text{if} & u(h_{(n)}, w_{(n)}h_{(n)} + \mu) & > & u(h_{(j)}, w_{(j)}h_{(j)} + \mu) \\
 & & & & j = 1, ..., n-1 \\
 & \text{and} & u(h_{(n)}, w_{(n)}h_{(n)} + \mu) & > & u_0
\end{array} \tag{4.15}$$

The observed job is the result of any of these n possibilities and therefore, the likelihood contribution of the observed job equals n times the probability that there are $n-1$ job offers with a utility level that does not exceed the utility level of the observed job, times the wage offer density function evaluated in the observed wage rate w_*, times the probability p_{l_*} of drawing the observed number of working hours h_{l_*}. The likelihood contribution of an observed wage-hours pair, conditional on e and the number of drawings n, becomes:

$$l(w_*, h_{l_*}|e, n) = n[P(w_*, h_{l_*}|e)]^{n-1}k(w_*, \eta' x, \sigma_v)p_{l_*} \qquad \begin{array}{r} l_* \in \{1, ..., m\} \\ u(h_{l_*}, w_* h_{l_*} + \mu) > u_0 \end{array} \tag{4.16}$$

where $k(.)$ is the log-normal density function of wage offers. Note that if n equals zero the likelihood contribution of the observed value is zero, as observing a job is in contradiction with the occurrence of zero job offers. If $n = 1$, there is no choice among different jobs and the likelihood contribution of observing (w_*, h_{l_*}) becomes just the job offer density evaluated in the observed job.

(4.16) is multiplied by the probability that n occurs, after which we sum over all n to obtain the likelihood contribution of the observed wage-hours package, conditional on the unobserved preference parameter e:

$$l(w_*, h_{l_*}|e) = \lambda \exp\{-\lambda[1 - P(w_*, h_{l_*}|e)]\} k(w_*, \eta' x, \sigma_v) p_{l_*} \qquad \begin{array}{r} l_* \in \{1, ..., m\} \\ u(h_{l_*}, w_* h_{l_*} + \mu) > u_0 \end{array} \tag{4.17}$$

For an individual who is not working, none of the n job offers generate a utility level which exceeds the utility level of not working, where we have to take into account that n actually may be zero. Then the probability that none of the n job offers is acceptable is given by

$$P(h = 0|e, n) = [P(0|e)]^n \tag{4.18}$$

where

$$P(0|e) = \sum_{l=1}^{m} p_l P(w \in A_{l0}(e)|e) \tag{4.19}$$

where $A_{l0}(e)$ is defined as in (4.10), with h_{l_*} replaced by zero and $g_l(e)$ in (4.12) and (4.13) is replaced by $g_{l0}(e)$, which is $g_l(e)$ with h_{l_*} replaced by zero.

Multiplying by the probability that n job offers arrive and summing over all possible n, including $n = 0$, gives the likelihood contribution of a non-working individual, conditional on e:

$$l(h = 0|e) = \exp\{-\lambda[1 - P(0|e)]\} \tag{4.20}$$

To remove the conditioning on the random preference parameter, the likelihood contribution has to be integrated over all $e, -\infty < e < \infty$. For the working individuals the likelihood contribution is zero if $u(h_{l_*}, w_* h_{l_*} + \mu) < u_0$ and therefore the effective integration region becomes

$$B := \{e | u_0 \leq u(h_{l_*}, w_* h_{l_*} + \mu)\} \tag{4.21}$$

The final likelihood contribution for an individual with a job becomes

$$l(w_*, h_{l_*}) = \int_B l(w_*, h_{l_*}|e) \frac{1}{\sigma_e} \phi\left(\frac{e}{\sigma_e}\right) de, l_* \in \{1, ..., m\}, 0 < w < \infty \tag{4.22}$$

where $\phi(.)$ is the standard normal density function. Note, that the range of w is $(0, \infty)$ after having integrated out e as the region B is non-empty for every positive wage rate, i.e. there always exists a range of random preferences such that working is preferred over non-working for every positive wage rate.

For the non-working individuals the likelihood contribution becomes

$$l(h = 0) = \int \exp\{-\lambda[1 - P(0|e)]\} \frac{1}{\sigma_e} \phi\left(\frac{e}{\sigma_e}\right) de \tag{4.23}$$

If the tax system is introduced the procedure remains basically the same. The probabilities in (4.12) have to be adapted and split up in accordance with the brackets in the tax system.

4.3 Estimation results

The model is estimated using a sample of 849 married women in the year 1985, obtained from the Organization for Strategic Labourmarket Research (OSA). In order to estimate the model the m hours categories of the hours offer distribution have to be chosen. To specify the discrete hours offer distribution, the hours are grouped into categories each of which contain four hours levels. As a consequence, the discrete hours distribution becomes

$$P(h = h_l) = p_l \text{ with } h_l = 4 \times l, l = 1, ..., m \tag{4.24}$$

In order to be able to identify all the probabilities, some equality restrictions are placed on probabilities of hours categories which have a low sample frequency. These restrictions are

$$\begin{aligned} p_1 = p_2 = p_3 = p_4 \\ p_6 = p_7 \\ p_{12} = p_{13} = p_{14} = p_{15} \end{aligned} \tag{4.25}$$

The value of m is chosen to be 15 in which case the maximum number of hours with a positive probability is 60, which coincides with the largest number of hours observed in the sample. The vector of individual characteristics X which appears in the utility function (4.8) consists of a constant (X_1 with parameter δ_1), the logarithm of the family size (X_2 with parameter δ_2) and a dummy indicator for the number of children with age below 6 (X_3 with parameter δ_3). The latter two are characteristics of which it is reasonable to assume that they affect the participation decision through the preferences, i.e. they affect the reservation utility level. The vector of characteristics x in the wage offer distribution consists of age variables to approximate the age-earnings profile and of education dummies as an approximation for human capital. To be more precise, x_1 is the constant term with parameter η_1, x_2 and x_3 are the logarithm of age/17 and its square, respectively, with paremeters η_2 and η_3, where the division by 17 is just a matter of normalization, and x_4, x_5, x_6, x_7 with parameters $\eta_4 - \eta_7$ are education dummies, with x_4 the lowest level of education, x_5 is the next to the lowest level etc.

As a point of departure the model is estimated with a linear budget constraint, whereas the parameter λ, representing the average number of job offers according to the Poisson distribution, does not depend on individual characteristics, which is also the case in the studies by Tummers and Woittiez (1991) and Van Soest, Woittiez and Kapteyn (1990). Table 4.1 shows the estimation results as well as estimates of the standard errors. The parameter estimates of log-family size and the number of children with age below six, $\hat{\delta}_2$ and $\hat{\delta}_3$ respectively, have a strong positive effect on the reservation utility level. The estimate $\hat{\lambda} = 36.7$ of the Poisson distribution seems to be rather high, indicating that the individuals in the sample are not that restricted. To compare, in Tummers and Woittiez (1991), the fixed maximum number of job offers in their binomial distribution was set equal to 10. The estimate has a sizeable standard error, however. From the estimates of the hours offer probabilities it can be seen that there are peaks at the numbers of hours of 20, 32 and 40, which can also be found back in the empirical distribution of labour supply. The age-earnings profile takes on its maximum value at the age of 33.

Having obtained parameter estimates it is possible to simulate the distribution of hours. The simulated hours frequencies can be compared with the observed hours frequencies. For each individual, a random preference parameter e and a number of job offers n is drawn from their assumed distributions. Then n wage-hours pairs are drawn, the utility levels are calculated and the highest utility level is compared with the utility of non-working to make the participation decision. This procedure is repeated 10 times and the resulting frequencies can be found in table 4.2. The second column in table 4.2 shows the observed frequencies and the third column shows the simulated frequencies. The participation decision is predicted well and the peaks at 20, 32 and 40 hours a week are predicted by the model.

Given the values of the parameter estimates it is possible to simulate the desired number of working hours, i.e. the number of working hours the individual would have chosen if she were not restricted in hours, which is the tangency point of the indifference curve and the budget constraint. Then we can simulate the frequency distribution of desired hours. In the case of a linear budget constraint the utility maximizing number of hours of utility function (4.8) is h with

$$h = 0 \quad \text{if } h^* \leq 0 \tag{4.26}$$

$$= h^* \quad \text{if } h^* > 0 \tag{4.27}$$

$$h^* = \beta\mu + \gamma w + X\delta + e \tag{4.28}$$

The simulation procedure is as follows. Draw a random preference parameter e and a number of job offers n. As the individual is not restricted in her working hours the only characteristic of a job that counts is the wage rate. Draw n wage rates and choose the highest. Calculate h^* in (4.28) and make the participation decision according to (4.26) and (4.27). Count the frequencies at which the hours occur. The results are in the fourth column of table 4.2. We can see that desired participation is somewhat higher than the actual participation: the frequency of observed participation is 0.390, whereas the frequency of desired participation is 0.473. This suggests that there is involuntary unemployment. Also, we see that the desired participation at 40 hours a week is about three times smaller than the actual participation at 40 hours a week. For positive hours, the hours distributions are plotted in the figures 4.1 and 4.2. In figure 4.1 we see the sample distribution and the distribution of simulated hours. In figure 4.2 the sample distribution can be compared with the distribution of desired hours. The distribution of desired hours clearly does not match the sample distribution. The peaks at the values of 20, 32 and 40 are not present in the distribution of desired hours.

So far, the parameter λ of the Poisson distribution which influences the number of job offers received by individuals has been the same for everybody. However, there are good reasons to assume that the number of job offers received by the individuals may differ across individuals. Young persons may get more job offers than older persons. The number of job offers may be different for workers who work in different sections of the economy. Different levels of education imply different kinds of jobs and the hiring procedures for the higher educated are usually not the same as those for individuals with a low level of education. Therefore, the Poisson parameter has been made dependent on individual characteristics:

$$\lambda_i = \exp(\theta' z_i), i = 1, ..., N \tag{4.29}$$

TABLE 4.1 PARAMETER ESTIMATES

Parameters	Estimates	Standard errors
β	-0.0287	0.00688
γ	6.473	1.605
δ_1 (const)	18.936	9.561
δ_2 (log fs)	-56.289	10.447
δ_3 (# child<6)	-27.231	7.386
σ_e	32.616	5.246
λ	36.662	32.260
σ_v	0.460	0.0596
η_1 (const)	1.445	0.290
η_2 (log(age/17))	1.368	0.236
η_3 $(\log(\text{age}/17))^2$	-1.019	0.169
η_4 (educ1)	-0.484	0.0699
η_5 (educ2)	-0.451	0.0685
η_6 (educ3)	-0.378	0.0624
η_7 (educ4)	-0.116	0.0678
$p_1 = p_2 = p_3 = p_4$	0.0743	0.00950
p_5	0.137	0.0202
$p_6 = p_7$	0.0511	0.00848
p_8	0.117	0.0177
p_9	0.0602	0.0134
p_{10}	0.232	0.0283
p_{11}	0.0323	0.0102
$p_{12} = p_{13} = p_{14} = p_{15}$	0.00556	0.00242

Log-likelihood value: -2002.4699

TABLE 4.2 SIMULATED HOURS FREQUENCIES

hours	Empirical frequencies	Simulated frequencies	Desired frequencies
0 (i.e. non-working)	0.610	0.609	0.527
4	0.0153	0.0200	0.0321
8	0.0318	0.0245	0.0316
12	0.0236	0.0232	0.0311
16	0.0259	0.0249	0.0306
20	0.0518	0.0502	0.0294
24	0.0318	0.0218	0.0283
28	0.00942	0.0235	0.0270
32	0.0495	0.0515	0.0255
36	0.0259	0.0253	0.0242
40	0.101	0.106	0.0227
44	0.0141	0.00132	0.0213
48	0.00353	0.00165	0.0197
52	0.00471	0.00177	0.0179
56	0.000	0.00188	0.0164
60	0.00118	0.00200	0.0147
> 60			0.100

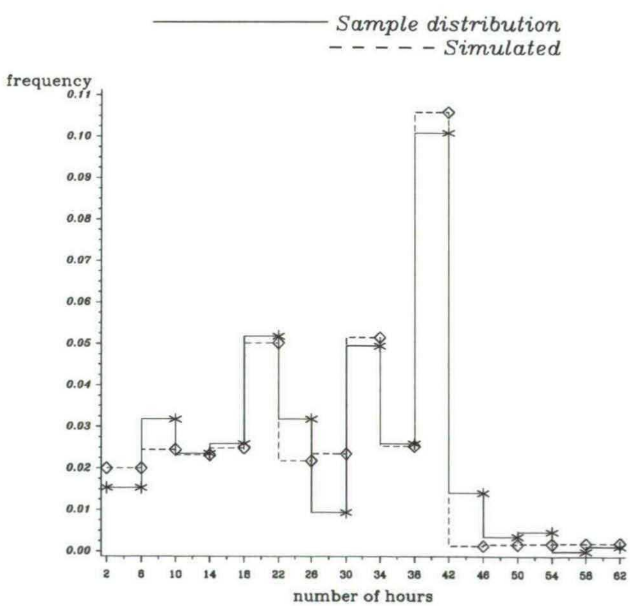

Figure 4.1: Distribution of working hours per week

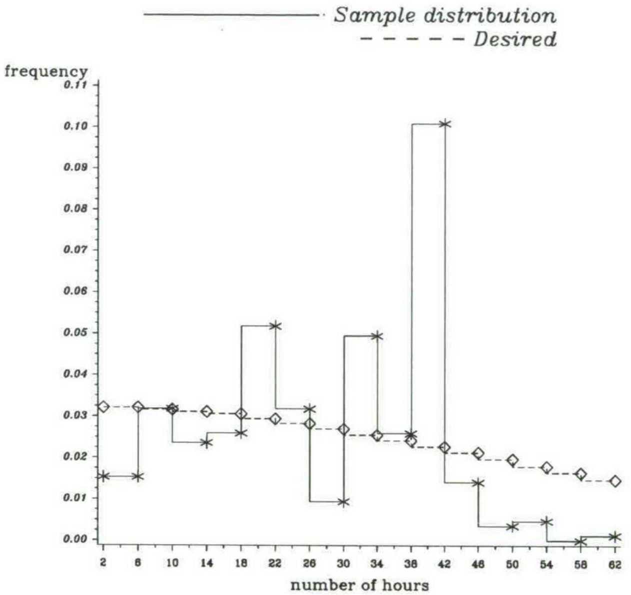

Figure 4.2: Distribution of working hours per week

The vector z_i contains the individual characteristics. The characteristics which are included are the constant term with parameter θ_1, an age variable with parameter θ_2, dummy variables which indicate the type of education (θ_3 and θ_4) and the level of education (parameters $\theta_5 - \theta_8$). Three types of education are distinguished: non-technical and non-commercial type of education, indicated by sec1 in table 4.3 (parameter θ_3), semi-technical and semi-commercial type of education (sec2, θ_4), and technical and commercial type of education for which no dummy variable is included. The estimation results are in table 4.3. Before considering the parameter estimates, the values of the log-likelihood functions of the specifications with and without a characteristic dependent Poisson parameter are compared. The value of the likelihood ratio test statistic to test the nullhypothesis that λ is independent of characteristics ($\theta_2 = \theta_3 = ... = \theta_8 = 0$) is 83.58, which is well above the critical value at the 5% level of 14.07.

Looking at the estimates of the utility parameters it can be seen that there are large increases in the (absolute) values of the estimates, including that of the standard deviation of the random taste parameter, as compared to the estimates of the invariant Poisson parameter model in table 4.1. Comparing the estimates of the parameters in θ with their standard errors we see that the age variable has a negative and significant effect, the dummy indicators for the semi-technical and semi-commercial type of education are positive and significant. The signs of the education dummy parameters are negative which indicates that the highest level of education has a higher Poisson parameter than the lower four levels of education. However, only the dummy for the second level is significant.

The distribution of hours is simulated in order to see how well the model tracks

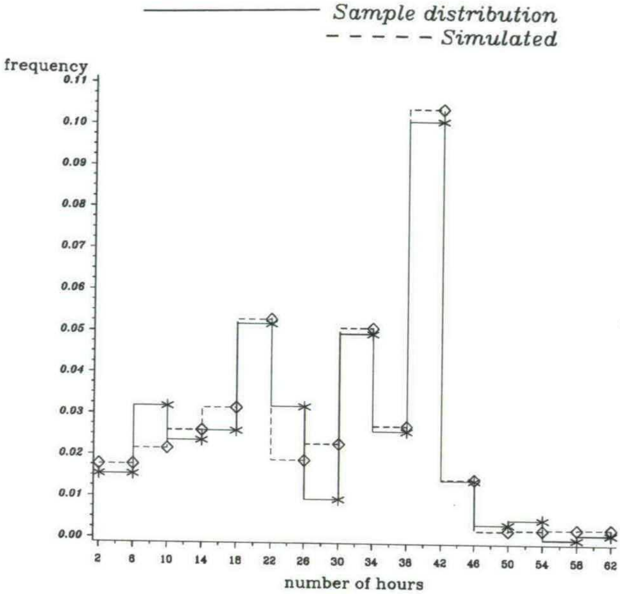

Figure 4.3: Distribution of working hours per week, variation in λ

the empirical distribution. The simulated frequencies are in table 4.4. The results are comparable to those in table 4.2. In the last column of table 4.4 are the desired frequencies, i.e. the frequencies of the number of hours which would have been chosen according to the estimates of the utility parameters if the individual were not restricted in the choice of hours. There is a large proportion of agents which is willing to work more than 60 hours a week. This is largely the result of the large variance of the unobserved taste parameter. Again, the hours distributions are plotted. Figure 4.3 shows the sample distribution and the distribution of simulated hours and figure 4.4 shows the sample distribution along with the distribution of desired hours. The distribution of simulated hours matches the sample distribution very well. The distribution of desired hours is so flat that it almost appears as a straight line in the figure.

So far, the following conclusions can be drawn. From the significance of the likelihood ratio test statistic it becomes clear that dependence of the distribution of the number of job offers on individual characteristics cannot be ignored, as was done in Tummers and Woittiez (1991) and Van Soest, Woittiez and Kapteyn (1990). At the same time, however, we see that if the parameter λ is made dependent on relevant individual characteristics, preferences seem to play no further role. The distribution of desired hours becomes very flat. This may of course be the result of the fact that demand side restrictions play such an important role on the labour market that they fully determine the behaviour of individuals. A different explanation for the phenomenon is that once the Poisson distribution has been made dependent on individual characteristics, there is simply not enough information in the data set to trace down the underlying preference structure. Furthermore, although it is possible to interprete the parameter estimates

TABLE 4.3 PARAMETER ESTIMATES WITH VARIATION IN λ

Parameters	Estimates	Standard errors
β	−0.163	0.319
γ	22.613	47.508
δ_1 (const)	409.473	739.121
δ_2 (log fs)	−421.260	820.003
δ_3 (# child<6)	−264.451	523.084
σ_e	248.687	484.629
θ_1 (const)	3.759	0.545
θ_2 (log(age/17))	−3.378	0.317
θ_3 (sec1)	1.417	1.427
θ_4 (sec2)	0.397	0.165
θ_5 (educ1)	−0.568	0.525
θ_6 (educ2)	−0.970	0.515
θ_7 (educ3)	−0.445	0.483
θ_8 (educ4)	−0.119	0.506
σ_v	0.265	0.00113
η_1 (const)	1.857	0.119
η_2 (log(age/17))	1.912	0.259
η_3 $(\log(age/17))^2$	−1.060	0.197
η_4 (educ1)	−0.397	0.0889
η_5 (educ2)	−0.288	0.0839
η_6 (educ3)	−0.279	0.0764
η_7 (educ4)	−0.0962	0.0824
$p_1 = p_2 = p_3 = p_4$	0.117	0.0120
p_5	0.162	0.0244
$p_6 = p_7$	0.0477	0.00857
p_8	0.0888	0.0154
p_9	0.0386	0.00918
p_{10}	0.125	0.0213
p_{11}	0.0148	0.00495
$p_{12} = p_{13} = p_{14} = p_{15}$	0.00194	0.000802

Log-likelihood value: -1960.6767

TABLE 4.4 SIMULATED HOURS FREQUENCIES

hours	Empirical frequencies	Simulated frequencies	Desired frequencies
0 (i.e. non-working)	0.610	0.603	0.484
4	0.0153	0.0177	0.00441
8	0.0318	0.0216	0.00443
12	0.0236	0.0260	0.00434
16	0.0259	0.0315	0.00432
20	0.0518	0.0529	0.00427
24	0.0318	0.0188	0.00437
28	0.00942	0.0228	0.00412
32	0.0495	0.0509	0.00439
36	0.0259	0.0272	0.00451
40	0.101	0.104	0.00426
44	0.0141	0.0144	0.00432
48	0.00353	0.00206	0.00432
52	0.00471	0.00221	0.00438
56	0.000	0.00241	0.00441
60	0.00118	0.00252	0.00446
> 60			0.450

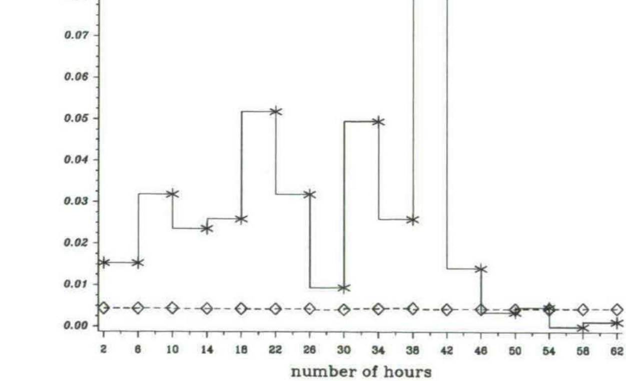

Figure 4.4: Distribution of working hours per week, variation in λ

of the characteristics in λ qualitatively, it is difficult to give an interpretation of their value. In the first model for example, the estimates of which are in table 4.1, we found an estimate of λ of about 37 which, in the context of this model and ignoring the standard error for the moment, would mean that each individual on average would have obtained 37 job offers. But as the model is static, the interpretation of this number is unclear as it has no time dimension. This problem would be solved if a dynamic model of sequential search would be formulated, in which data on unemployment duration would provide additional information, both to estimate and to interpret the parameters of the number of job offer distribution and the parameters of the utility function. This approach is followed in chapter 5 of this thesis.

We now drop the assumption that the budget constraint is linear. A progressive tax system causes the net wage rate and the number of hours to be negatively related in the budget constraint. Although a fully structural treatment of a system of labour income taxes is in principle implementable in the model, we abstain from it here because of its rather heavy computational burden. Moffit (1984) proposed to formulate the wage rate as a second order polynomial in the number of working hours. The advantage of this approach, over the introduction of a fully specified tax system, is that we can actually test for the dependence of net wages on the number of working hours.

The wage equation is extended by introducing hours.

$$\ln w = \eta' x + \tau_1 h + \tau_2 h^2 + v \tag{4.30}$$

where w is the net wage rate, x is a vector of individual characteristics and v is again a normally distributed random variable. The wage equation in (4.3) has been extended with a term which is linear in the number of working hours and a term which is quadratic in the number of working hours.

There is another reason why wages may depend on hours which is interesting to mention in this context. At institutionally determined numbers of hours like 20, 32 and 40, the offered wages may be higher for given individual characteristics and therefore hours restrictions need not be the only explanation for observing peak levels. To capture this possible relation between wages and hours in the model, dummy variables could be taken up in the wage equation. However, inspection of the data led us to abstain from it because no such relation appears to be present. Therefore, we shall assume from now on that the hours terms in the wage equation represent the tax system. The gross wage rate, w_G, is given by

$$w_G = \exp(\eta' x + v) \tag{4.31}$$

and the relation between the gross and the net wage rate is

$$w = (1 - \tau(h)) w_G \tag{4.32}$$

where $\tau(h)$ is the marginal tax rate which should be between zero and one and which is non-increasing in h for a given wage rate if the tax system is progressive. From (4.30), (4.31) and (4.32) we derive:

$$0 < \tau(h) < 1 \quad \text{if} \quad \tau_1 h + \tau_2 h^2 < 0 \tag{4.33}$$

$$\tau'(h) < 0 \quad \text{if} \quad \tau_1 + 2\tau_2 h < 0 \tag{4.34}$$

From the parameter estimates it can be checked whether these conditions are satisfied, i.e. it can be checked whether our assumption, that the hours terms in the wage equation mainly represent the tax system, is fulfilled.

The parameter estimates are given in table 4.5. If the estimates of the utility parameters are compared with the estimates of the basic model in table 4.1, it can be noticed that in the present model preferences play a more pronounced role. The estimate of the standard deviation of the random preference parameter has decreased substantially. The parameter estimate of the number of job offers is also reduced. There is a striking difference in the estimates of the hours offer probabilities, especially of those for the hours categories above 40 hours a week. In the basic model, there was a close relation between the observed frequencies of hours and these probabilities. The explanation for the low frequency of hours above 40 was that these values of hours are hardly offered. In table 4.5 it can be seen that according to the present model the probability of receiving hours levels of 44 or higher is not that low at all. However, the marginal increase in income as a result of working an additional hour is apparently so low, as compared to the effect on the marginal utility of leisure, that the individuals are in general not willing to supply these high hours levels.

Looking at the estimates of τ_1 and τ_2 it can be observed that τ_1 is positive but insignificant and that τ_2 is significantly negative. Checking relations (4.33) and (4.34) it follows that (4.33) is satisfied if $h > 5.7$ and relation (4.34) is satisfied for $h > 2.9$. This, together with the insignificance of τ_1, is in accordance with our interpretation of the hours terms.

The likelihood ratio test statistic to test the hypothesis $\tau_1 = \tau_2 = 0$ has the value 16.596. The critical value at the 5% level is 5.991, which means that the hypothesis is rejected.

The simulation results in table 4.2, 4.4 and 4.6 and the graphs in the figures 4.1, 4.2 and 4.3 provide an informal way of testing the model. To formalize the testing of the model the chi-square test statistic of Andrews (1988) can be calculated. As the aim of the testing procedure is to test the predictive power of the model, only the values of the endogenous hours variable are categorized into cells. Andrews' test statistic is calculated using three different partitions of the hours variable. Partition 1 is the most refined and coincides with the catogorisation of the hours in the simulation studies in tables 4.2, 4.4 and 4.6. In partition 2 the values of the hours have been classified according to the restrictions which have been imposed on the probabilities of the hours offer distribution. Partition 3 classifies hours in only two different groups i.e. positive and non-positive hours. The test statistic calculated with partition 3 can be used to test the predictive power of the model with respect to the participation decision. The values of the test statistic are in table 4.7. It is clear that all of the three model specifications are rejected by the three test statistics.

Estimation experiment with both an hours dependent wage equation and an individual specific Poisson parameter gave similar results as the results in table 4.3, i.e. a flat utility function.

TABLE 4.5 PARAMETER ESTIMATES, MODEL WITH HOURS DEPENDENT WAGES

Parameters	Estimates	Standard errors
β	-0.0172	0.00476
γ	4.185	1.201
δ_1 (const)	14.670	5.664
δ_2 (log fs)	-31.568	7.887
δ_3 (# child<6)	-15.973	5.077
σ_e	16.675	4.458
λ	17.578	10.952
σ_v	0.356	0.0362
η_1 (const)	2.218	0.187
η_2 (log(age/17))	1.388	0.242
η_3 $(\log(\text{age}/17))^2$	-1.388	0.173
η_4 (educ1)	-0.495	0.0750
η_5 (educ2)	-0.460	0.0719
η_6 (educ3)	-0.359	0.0674
η_7 (educ4)	-0.125	0.0721
τ_1	0.00208	0.00652
τ_2	-0.000362	0.000122
$p_1 = p_2 = p_3 = p_4$	0.00760	0.00472
p_5	0.0210	0.0120
$p_6 = p_7$	0.0121	0.00656
p_8	0.0504	0.0235
p_9	0.0433	0.0184
p_{10}	0.312	0.0952
p_{11}	0.0923	0.0299
$p_{12} = p_{13} = p_{14} = p_{15}$	0.107	0.0451

Log-likelihood value: -1985.8735

TABLE 4.6 SIMULATED HOURS FREQUENCIES, HOURS DEPENDENT WAGES

hours	Empirical frequencies	Simulated frequencies	Desired frequencies
0 (i.e. non-working)	0.610	0.600	0.137
4	0.0153	0.0251	0.0382
8	0.0318	0.0257	0.0498
12	0.0236	0.0250	0.0627
16	0.0259	0.0227	0.0860
20	0.0518	0.0537	0.102
24	0.0318	0.0247	0.107
28	0.00942	0.0183	0.113
32	0.0495	0.0514	0.102
36	0.0259	0.0269	0.0999
40	0.101	0.103	0.0729
44	0.0141	0.00142	0.0219
48	0.00353	0.00640	0.00612
52	0.00471	0.00199	0.00188
56	0.000	0.000599	0.000234
60	0.00118	0.000110	0.000
> 60			0.000

TABLE 4.7 ANDREWS' CHI-SQUARE TEST STATISTIC

chi-square statistic	partition 1	partition 2	partition 3
fixed λ model	395.460	47.466	33.534
variable λ model	269.233	39.481	19.062
hours dependent wages	216.504	49.494	34.337
critical value at 5% level	24.996	15.507	3.841

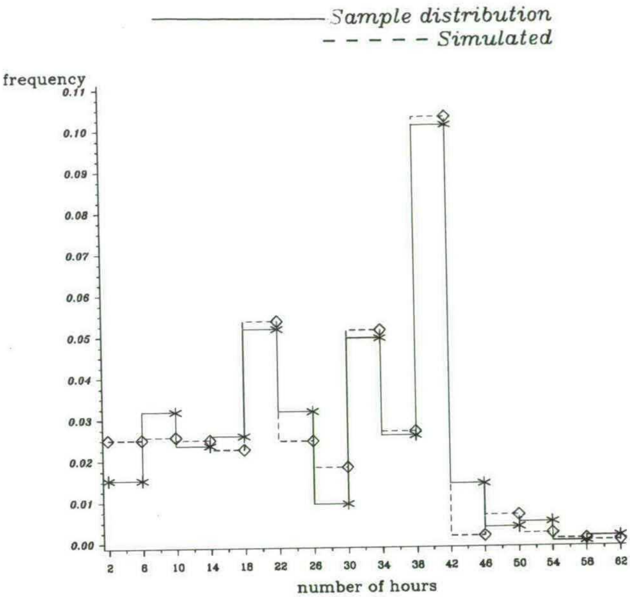

Figure 4.5: Distribution of working hours per week, hours dependent wages

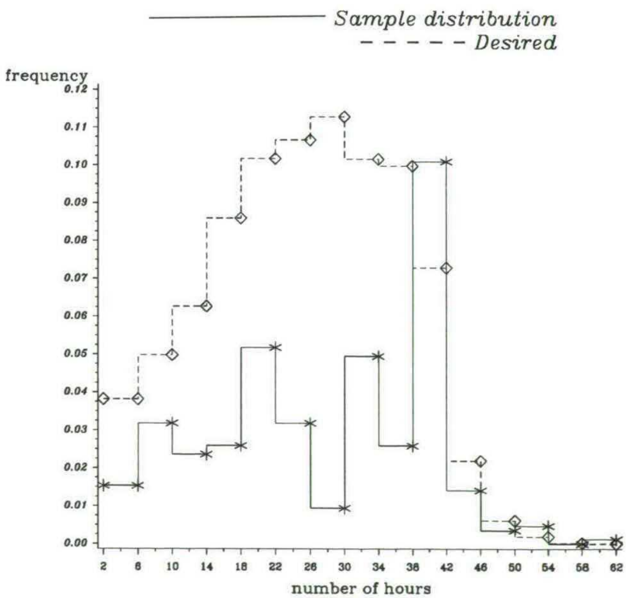

Figure 4.6: Distribution of working hours per week, hours dependent wages

4.4 Conclusions

In this chapter a static model of labour supply with job offer restrictions has been formulated. A job, as it is offered by an employer to an individual, consists of both an hours component and a wage component, thereby linking the standard job search model with the static model of labour supply with hours restrictions. Because of the structure of the model, the parameters of the wage distribution and the hours distribution have been estimated simultaneously. The number of job offers has been assumed to be Poisson distributed and the parameter of the Poisson distribution has been made dependent on relevant individual characteristics to examine their effect on the average number of job offers. The results of the likelihood ratio test indicate the significance of the set of individual characteristics in the Poisson distribution. The result of the introduction of the individual characteristics in the Poisson distribution is that the parameters of the underlying utility specification cannot be traced down anymore. That is, the utility function becomes a flat and approximately uniform random function. This suggests that the available data provide too little information to obtain sensible estimates of the utility parameters. Although the simulated hours distribution in this type of model fits the labour supply data better than in the neoclassical model, as argued in Van Soest, Woittiez and Kapteyn (1990), this is mainly the result of the discrete specification of the hours offer distribution.

We have pleaded for the introduction of a time dimension into the model by formulating a sequential job search model in which duration data provide additional information in the estimation and interpretation of the model parameters. This idea is worked out in chapter 5 of this thesis.

Estimation results of a model specification in which the wage equation contains the number of weekly working hours indicate the significance of the presence of working hours in the equation. Moreover, the estimation results are consistent with the interpretation that working hours arise in the wage equation as a result of the tax system. The estimates of this specification suggest that the fact that empirical frequencies of weekly working hours above 40 are low, is not mainly the result of low offer probabilities, as was the case in previous specifications, but is the consequence of a low marginal increase in income of working an additional hour as compared to the marginal utility of leisure at high hours levels.

Simulation experiments in which the simulated hours distribution, as generated by the model, was compared to the empirical hours distribution show that the model predicts participation and various peaks in the hours distribution well. However, a formal testing procedure rejects the hypothesis that the simulated data and the observed data have the same distribution.

Chapter 5

Job search theory, labour supply and unemployment duration

5.1 Introduction

This chapter presents models in which elements of job search theory and the labour supply literature are combined. A functional form for the models will be specified and the functional form will be estimated structurally. Flinn and Heckman (1982a) present an overview of the estimation of structural job search models. Applications can be found in Narendranathan and Nickell (1985) and Van den Berg (1990a). In order to find closed form solutions for the model, restrictive assumptions have to be made and this is the reason why the estimation of structural job search models has not become as popular as the reduced form duration models, see Flinn and Heckman (1982b) and Kiefer (1988) for an overview.

Nevertheless, there is still room for extension in the specification of structural job search models as compared with the models which have been estimated up till now. Notably, the assumption that the wage rate is the only component in the objective function of the individual decision maker can be relaxed. The standard job search framework is only concerned with the choice of a job on the basis of the wage rate. On the other hand, there exists an extensive literature on labour supply models. In these models the availability of a job is given and the emphasis is on the participation decision and the choice of the number of hours. Until now few attempts have been made to integrate these two types of models. The aim of this chapter is to extend the standard job search framework with elements of labour supply theory.

Basically, two routes can be followed. The first one is to make the neo-classical assumption that individuals can choose their labour supply optimally, given the level of their wage rate. In this case, the assumption from standard job search theory, that the wage rate is the only job component on which the job acceptance decision is based, remains valid. The neo-classical assumption is usually not supported by empirical evidence on the distribution of hours worked, where we typically see peaks at various levels of hours, e.g. at 40 hours a week. Therefore, the second possibility is to assume that hours are, just like the wage rate, a component of the job offer. The first attempt to introduce hours in a job search context was made by Kiefer (1984,1987). Blau (1991) estimated a search model in which both wages and hours appear in the utility function. In the labour supply literature restrictions on hours were introduced in a static model by Dickens and Lundberg (1985). Their model was further developed by Tummers and Woittiez (1991) and Van Soest, Woittiez and Kapteyn (1990). A different route, to handle hours restrictions, was followed by Rettore (1990). In the previous chapter a static model with hours restrictions was formulated which can be interpreted as a static version of a dynamic job search model in which the rate of time preference is infinite.

In this chapter, specific attention is paid to the stochastic specification. A random preference term is included in the utility function. As a result, the reservation wage rate will be random as well, as opposed to earlier work, e.g. Van den Berg (1990a) and Narendranathan and Nickell (1985), in which the reservation wage rate could only be changed by a change in the model parameters. Both the model with neo-classical labour supply and the model with hours restrictions will be considered in this chapter. Data on unemployment duration and post unemployment job characteristics are used to estimate the parameters of the utility function, the parameters of the job offer distribution and the job offer arrival rate. The likelihood function can be formulated, conditional on the random preferences, after which the random preferences can be integrated out. This integration procedure can be costly if it has to be performed numerically, which is definitely the case here, because the integrand contains the reservation wage rate which is the solution of a fixed point problem. To save computing time we can make use of simulation estimators. McFadden (1989) introduced a simulation estimator which is consistent for a fixed number of simulation replications in the context of a multinomial choice model. In chapter 3 this method was adapted to the limited dependent variables model and applied to the neo-classical labour supply model. In the application in chapter 3 use is made of simulation methods to integrate out an unobserved random preference variable. Various simulation estimators in the context of models with unobserved random variables and the specific problems which actually arise in this context will be discussed.

In section 5.2 the model without hours restrictions will be set up. First, attention will be paid to the assumptions that have to be made in order to be able to estimate the model. After that the reservation wage equation will be derived which specifies the strategy followed by the individual. The likelihood function is specified after which it is indicated how the likelihood estimator can be replaced by a simulation estimator. Thirdly, specific functional forms for the model will be chosen. Fourth, the available dataset is discussed. Finally, the model will be estimated and estimation results will be presented. In section 5.3 the neo-classical assumption of no hours restrictions is dropped and a job offer is supposed to consist of two characteristics, i.e. the number of hours and the wage rate. Instead of a single reservation wage rate there now exists a unique reservation utility level. All job offers which yield a higher utility value than this reservation utility value are acceptable. Again, the model will be estimated and estimation results will be presented. Section 5.4 presents residual analysis and the final section concludes.

5.2 Job search without hours restrictions

5.2.1 The model

In this section a job search model is presented in which an unemployed individual maximizes the discounted sum of expected future utility flows, under the assumption that he knows the probability distribution according to which job offers arrive. The utility flow is a function of income and labour supply and it is assumed that once the wage rate is known, hours can be chosen optimally by maximizing the utility function subject to the budget constraint. Implicitly, the assumption of no hours restrictions is also made in the standard job search model in which the individual just maximizes the discounted sum of expected future utility flows. Therefore, the value of work function in the standard job search framework can be interpreted as the discounted sum of indirect utility flows. In the next section the neo-classical assumptions that individuals can choose hours optimally will be dropped. Because the neo-classical assumption elucidates the relation between the present model and the standard job search framework and uncovers some of its implicit assumptions it is a good point of departure.

In order to be able to find closed form solutions for the model, some possibly restrictive assumptions have to be made. Most of these assumptions are standard assumptions which are usually made in structural models of job search. They can be found in Mortensen (1986). We deviate from the standard framework in Mortensen (1986) by including labour supply in the utility function. A different approach of incorporating labour supply in the context of a job search model can be found in Burdett and Mortensen (1978).

The assumptions are the following:

1. The individual maximizes a discounted sum of future expected utility flows, subject to the budget constraint and the job offer process:

$$\max_{c_t,h_t} E\int_t^{\infty} u(y_s,h_s;\epsilon)e^{-\rho(s-t)}ds \tag{5.1}$$

   where y_t is income in period t, h_t is labour supply in period t, ρ is the discount rate, ϵ is an individual specific, time independent unobserved random taste parameter, known to the individual. The appearance of the expectation sign refers to the uncertainty about the future state, i.e. the uncertain job possibilities and the associated wage rates. The job offer arrival and wage processes are specified below. By the inclusion of labour supply in the utility function we deviate from the standard job search framework, in which the utility function contains income only.

2. The income consists of a state dependent component and a state independent component (non-labour income). If employed, income equals the sum of labour income and non-labour income:

$$y_t = wh_t + \mu \tag{5.2}$$

where w is the wage rate and μ is non-labour income. If unemployed, income equals the sum of the unemployment benefit payments b and the state independent income:

$$y_t = b + \mu \tag{5.3}$$

As a consequence, the level of the benefit payments will influence the job acceptance decision.

3. A job offer consists of a wage rate w. Job offers arrive randomly according to a Poisson process with parameter λ from a distribution function $F(.;\psi,\tau)$ with accompanying density function $f(.;\psi,\tau)$, with ψ the location parameter and τ the scale parameter. The distribution function is known to the individual. The domain of $F(.;\psi,\tau)$ is $(0,\infty)$.

4. The model is stationary, i.e. the job offer arrival rate λ, the unemployment benefit level b, the wage distribution $F(w;\psi,\tau)$ and the non-labour income μ are independent of both calendar time and elapsed duration.

5. Once the unemployed has accepted a job it will be kept forever.

6. The utility function has the properties:

$$\begin{aligned} \frac{\partial u}{\partial y} &> 0 && (5.4)\\ k(h) &:= u(wh+\mu, h;\epsilon) && (5.5)\\ k"(h) &< 0 \text{ for all } h > 0 && (5.6) \end{aligned}$$

Under these assumptions it can be derived that there exists a unique reservation wage $\xi(\epsilon)$. All wage offers above the reservation wage rate are acceptable, whereas those below will be rejected.

Assumption 3 is a standard assumption in the job search framework. Assumption 4 is an assumption which we need to arrive at a closed form solution. Without this assumption the reservation wage rate can only be defined implicitly in terms of a differential equation, which can only be solved if not too complicated assumptions for the process of exogenous variables are specified. Van den Berg (1990b) relaxes the stationarity assumptions by introducing general forms of non-stationarity. His empirical application, however, remains restricted to non-stationarity in the benefit level b, in which, after the unemployment spell has reached a specified length, a discrete jump takes place whereafter it remains constant. In the duration model literature, in which reduced form models are estimated, it turns out to be difficult to distinguish empirically between negative duration dependence and unobserved heterogeneity, i.e. once unobserved heterogeneity is introduced, the presence or absence of duration dependence is hard to establish. As we have included a random preference parameter in the utility function we hope to be able to catch the heterogeneity in the model.

Assumption 5 is made to be able to find a closed form expression for the value of a job, which then can be compared with the value of search. This assumption is of course restrictive if a job is taken on for a short period only, but if a fixed job is taken on it is not unreasonable to assume that the individual acts as if he will hold the job forever. Moreover, the general form of the reservation wage equation remains valid if it is assumed that there is a constant layoff rate (see e.g. Flinn and Heckman (1982a)). In that case we have to be careful in the interpretation of the rate of time preference ρ.

The final assumption on the form of the utility function is made to ensure that a higher wage level is always preferred to a lower, and to ensure that, once a wage level is chosen, within period utility will reach a maximum if it is maximized with respect to labour supply, subject to the linear budget constraint. Note, that no restrictions are placed on the sign of the derivative of the utility function with respect to labour supply. This is done because results of Narendranathan and Nickell (1985) and Van den Berg (1990c) indicate that individuals might value unemployment, i.e. total leisure, lower than having a job. However, their specification of the objective function was very straightforward in the sense that there is a discrete jump between the utility values of unemployment and having a job. The size of this jump was estimated by them with the result mentioned before. The restrictiveness of their specification and the fact that their estimated utility parameters were not based on the after spell income data directly, because the parameters of the wage equation were estimated separately, using some ad hoc truncation rule, imply that care must be taken in the interpretation of the results. Any acceptance of a job which cannot be fully explained by the individual characteristics in their model, can only be explained by the jump in the utility value, so the negative valuation of total leisure may have been only the result of their restrictive specification. Nevertheless, their results induce us not to restrict our utility function by the requirement that it is decreasing in labour supply everywhere. This implies that the reservation wage rate may become negative, in which case any job offer is acceptable to the individual.

The reservation wage $\xi(\epsilon)$ is implicitly defined by the following equation, the derivation of which can be found in appendix A:

$$\nu(\xi(\epsilon),\mu;\epsilon) = u(b+\mu,0;\epsilon) + \frac{\lambda}{\rho}\int_{\xi(\epsilon)}^{\infty} [\nu(w,\mu;\epsilon) - \nu(\xi(\epsilon),\mu;\epsilon)]\, dF(w;\psi,\tau) \tag{5.7}$$

in which $\nu(w,\mu;\epsilon)$ is the indirect utility function which is the result of substituting the labour supply function into the direct utility function. The indirect utility function is well-defined whenever the labour supply function is positive, i.e. whenever the wage rate w exceeds $w_0(\epsilon)$ which is defined by

$$\nu(w_0(\epsilon),\mu;\epsilon) = u(\mu,0;\epsilon) \tag{5.8}$$

$w_0(\epsilon)$ is the reservation wage rate of the static model without unemployment benefits ($\rho \to \infty, b = 0$). It can easily be shown that the dynamic reservation wage rate $\xi(\epsilon)$ defined by (5.7) always exceeds the static model reservation wage rate $w_0(\epsilon)$. This ensures that labour supply is positive whenever $w > \xi(\epsilon)$.

In words, the reservation wage rate is the wage rate which equates the value of the indirect utility function to the sum of the direct utility value of being unemployed and the expected gain of receiving wage offers in the future.

Under the above assumptions it is a general result that the distribution of completed spells of unemployment, conditional on ϵ, is exponential with escape rate $\theta(\epsilon)$, which loosely speaking equals the probability of getting an acceptable job offer.

$$\theta(\epsilon) = \lambda \bar{F}(\xi(\epsilon)) \tag{5.9}$$

The density function of a completed spell of unemployment, conditional on ϵ, is

$$k(t|\epsilon) = \theta(\epsilon)\exp\{-\theta(\epsilon)t\}, 0 < t < \infty \tag{5.10}$$

In the formulation of the likelihood function the sampling scheme has to be taken into account. In practice, usually two sampling methods are distinguished, i.e. sampling the flow and sampling the stock. In the flow sample the observation period starts at a certain point in time, whereafter, all individuals who become unemployed after the beginning of the observation period are drawn into the sample and their completed spells are observed. There may be right hand censoring because the unemployment spell has not ended before the end of the observation period. This type of sampling is the most straightforward because we can directly use the distribution of completed spells. In the stock sampling scheme we take a given point in time and sample individuals who are unemployed at that point in time from the stock of unemployed. We assume that we can observe both how long they have already been unemployed (i.e. the backward recurrence times) and how long they will be unemployed from the point of sampling on (the forward recurrence times), again with possible right hand censoring. Suppose, that the backward recurrence time is indicated by p and the forward recurrence time by r which implies that the total spell of unemployment t is the sum of p and r. Now, different routes in the formulation of the likelihood function can be followed. We can formulate the joint distribution of forward and backward recurrence times, in which case we either have to model the process of inflow in the state of unemployment, or have to make the assumption that the inflow rate is constant. The second possibility is to condition on the backward recurrence time in which case we do not have to make assumptions about the inflow rate. The latter procedure will be followed in this chapter. In the case without unobserved heterogeneity the backward recurrence times simply drop out because of the stationarity assumption.

For ease of exposition, the likelihood contribution for the flow sample will be derived first and it is supposed that the observation period, which starts with the beginning of the unemployment spell for individuals in the flow sample, is of length M.

During the observation period the individual can be observed to accept a job or not. If no job is accepted during the observation period, the only information we have is that the duration of the unemployment spell lasts longer than the observation period. Then the likelihood contribution of such an individual is given by

$$l_u(\eta|\epsilon) = \exp\{-\theta(\epsilon)M\} \tag{5.11}$$

where η is the vector of parameters. If ϵ has a density function $g(.;\sigma_\epsilon^2)$ then the conditioning on ϵ can be removed by simply integrating out ϵ. The unconditional likelihood contribution becomes

$$l_u(\eta) = \int_{-\infty}^{\infty} \exp\{-\theta(\epsilon)M\} g(\epsilon, \sigma_\epsilon^2) d\epsilon \tag{5.12}$$

For individuals who accept a job during the observation period we can distinguish between individuals whose after unemployment spell wage rate and hours are observed and individuals whose job characteristics are unobserved. The optimal labour supply h^* for an individual with wage rate w and non-labour income μ is given by the labour supply function

$$h^* = \bar{h}(w, \mu; \epsilon) \tag{5.13}$$

Recall, that optimal labour supply is always positive for wage rates exceeding the reservation wage rate. If we assume that observed labour supply h is measured with a multiplicative measurement error $\exp(v), -\infty < v < \infty$, the density of observed labour supply, conditional on the wage rate and the random taste parameter can be derived from the density of measurement error v, by making the transformation

$$h = \bar{h}(w, \mu; \epsilon)\exp(v) \tag{5.14}$$

Let the resulting density function be denoted by $r(h|w,\epsilon)$. If an individual is observed to be working at a wage w and hours h, this means that the observed wage rate must exceed the reservation wage rate $\xi(\epsilon)$, which means that the density of observed wages, conditional on ϵ, is truncated. The likelihood contribution of individuals who accepted a job and whose after spell wages and hours are observed and are equal to w and h respectively, and whose unemployment duration equals t, conditional on ϵ, is given by

$$\begin{aligned} l_{co}(\eta|\epsilon) &= \theta(\epsilon)\exp\{-\theta(\epsilon)t\}r(h|w,\epsilon)\frac{f(w)}{T(\epsilon)}, 0 < t < M, h > 0, w > T(\epsilon) && (5.15)\\ &= 0 \text{ otherwise} && (5.16)\end{aligned}$$

where $T(\epsilon)$ is the truncation probability which is defined by:

$$\begin{aligned} T(\epsilon) &= \bar{F}(\xi(\epsilon)) \text{ if } \xi(\epsilon) > 0 && (5.17)\\ &= 1 \text{ if } \xi(\epsilon) \leq 0 && (5.18)\end{aligned}$$

To remove the conditioning on ϵ we have to integrate over all values of ϵ for which $w > \xi(\epsilon)$. The unconditional likelihood contribution becomes:

$$l_{co}(\eta) = \int_{I_w} \theta(\epsilon)\exp\{-\theta(\epsilon)t\}r(h|w,\epsilon)\frac{f(w)}{T(\epsilon)}g(\epsilon,\sigma_\epsilon^2)d\epsilon, t > 0, h > 0, w > 0 \tag{5.19}$$

where

$$I_w = \{\epsilon|\xi(\epsilon) < w\} \tag{5.20}$$

For some of the individuals who are observed to accept a job during the observation period the after spell job characteristics may be unobserved. In that case the unobserved taste parameter will be integrated out and the likelihood contribution becomes

$$l_{cu}(\eta) = \int_{-\infty}^{\infty} \theta(\epsilon)\exp\{-\theta(\epsilon)t\}g(\epsilon,\sigma_\epsilon^2)d\epsilon, 0 < t < M \tag{5.21}$$

Now we return to the formulation of the likelihood contributions of the individuals in the stock sample. The point of right hand censoring M now is the length of the period which starts at the point of sampling and ends at the end of the observation period. As said before, we condition on the backward recurrence time p, which implies that we condition on duration t being longer than p. As a result, we simply have to divide the likelihood contribution derived above by the probability that t exceeds p, see e.g Ridder (1984). The implicit assumption on the inflow rate which is made in following this procedure is that the inflow rate into the state of unemployment does not depend on the unobserved random variable ϵ.

$$P(t > p) = \int_{-\infty}^{\infty} \exp\{-\theta(\epsilon)p\} g(\epsilon, \sigma_\epsilon^2) d\epsilon \tag{5.22}$$

where in (5.11) and (5.12) M has to be replaced by $p + M$ and the region for t in (5.15) and (5.21) becomes $p < t < p + M$.

5.2.2 Simulation estimators for models with unobserved heterogeneity

In this section two simulation estimators for models with unobserved random variables which need to be integrated out are discussed. The first method simulates the vector of scores of the log-likelihood function in such a way that the resulting simulated score vector has expectation zero at the true parameter vector. A drawback of the method is that random variables need to be drawn from the distribution of the unoberved random variable, conditional on the observed random variable. The denominator of the expression for the conditional density function of the unobserved random variable, conditional on the observed random variable, is the marginal density function of the observed variable and this density contains the integral whose evaluation we want to avoid by making use of simulation techniques. Drawing from e.g. the marginal distribution introduces a bias in the simulation of the vector of scores and as a consequence the resulting estimator will be inconsistent. The second method, to be considered below, is called smooth simulated maximum likelihood estimation (SSML) as described by Börsch-Supan and Hajivassiliou (1993).

The likelihood function presented in the previous section is of the type in which an unobserved random variable is integrated out. Let the unknown random variable be denoted by x with marginal density $f(x|\eta)$ and let the observed random variable, or vector of random variables be denoted by y, with density function conditional on x given by $\tilde{f}(y|x;\eta)$, in which η is the vector of parameters. The log-likelihood contribution of a single individual is given by

$$L(\eta|y) = \ln\left[\int_{-\infty}^{\infty} \tilde{f}(y|x;\eta) f(x;\eta) dx\right] \tag{5.23}$$

Suppose that by transformation the integral can be written as

$$L(\eta|y) = \ln\left[\int_{-\infty}^{\infty} f(y|u;\eta)\phi(u) du\right] \tag{5.24}$$

where $\phi(.)$ is a density function which does not depend on the parameters of interest. In the context of the previous section $d \ln l_u(\eta)$ has the form of $L(\eta|y)$, where d is a dummy variable taking on the value 1 for uncompleted spells and the value 0 for completed spells. Adding over all individuals the log-likelihood function becomes

$$L(\eta; y_1, ..., y_N) = \frac{1}{N}\sum_{i=1}^{N} \ln\left(\int_{-\infty}^{\infty} f(y_i|u;\eta)\phi(u)du\right) \tag{5.25}$$

assuming that the y_i are i.i.d. It is a well known property of the log-likelihood function that the expectation of the vector of scores equals zero at the true parameter vector η_0. It is this property that is exploited in the derivation of moment conditions. The vector of scores is:

$$\frac{1}{N}\frac{\partial L(\eta|y_1, ..., y_N)}{\partial \eta} = \frac{1}{N}\sum_{i=1}^{N} \frac{\int_{-\infty}^{\infty} \partial f(y_i|u;\eta)/\partial\eta\phi(u)du}{\int_{-\infty}^{\infty} f(y_i|u;\eta)\phi(u)du} \tag{5.26}$$

The problem is that the integral appears in both the numerator and the denominator and if we want to exploit the possibility of keeping the number of drawings R, used in the simulation, small, the simulator has to enter the moment conditions linearly, see, e.g. Gouriéroux and Montfort (1989). A simulator can be constructed in the following way. Draw R i.i.d. random numbers u_{ir} from $f(.|y_i)$, the density of u conditional on the observed value y_i, for every individual i, $i = 1, ..., N$, and use the moment functions $S(\eta; y_i, u_{i1}, ..., u_{iR}, i = 1, ..., N)$:

$$S(\eta; y_i, u_{i1}, ..., u_{iR}, i = 1, ..., N) = \frac{1}{N}\frac{1}{R}\sum_{i=1}^{N}\sum_{r=1}^{R} \frac{\partial f(y_i|u_{ir};\eta)/\partial\eta}{f(y_i|u_{ir};\eta)} \tag{5.27}$$

To show that $S(.)$ has the same expectation properties as the true vector of scores, we take the expectation with respect to the drawings u_{ir}, thereby conditioning on the $y_i, i =, ..., N$. The conditional distribution of u_{ir} given y_i is

$$f(u_{ir}|y_i) = \frac{f(y_i|u_{ir};\eta)\phi(u_{ir})}{\int_{-\infty}^{\infty} f(y_i|\tilde{u};\eta)\phi(\tilde{u})d\tilde{u}} \tag{5.28}$$

Taking expectations yields

$$E(S(.)|y_i, i = 1, ..., N) =$$

$$\frac{1}{N}\frac{1}{R}\sum_{i=1}^{N}\sum_{r=1}^{R}\int_{-\infty}^{\infty} \frac{\partial f(y_i|u_{ir};\eta)/\partial\eta\phi(u_{ir})}{\int_{-\infty}^{\infty} f(y_i|\tilde{u};\eta)\phi(\tilde{u})d\tilde{u}}du_{ir} = \frac{1}{N}\sum_{i=1}^{N}\frac{\int_{-\infty}^{\infty}\partial f(y_i|u;\eta)/\partial\eta\phi(u)du}{\int_{-\infty}^{\infty} f(y_i|\tilde{u};\eta)\phi(\tilde{u})d\tilde{u}} \tag{5.29}$$

which equals the true vector of scores. As a consequence, at the true parameter vector η_0 the simulated moments conditions have expectations zero. In (5.12) we simulate of course only one term of the score vector, i.e. the derivative of

$$\frac{1}{N}\sum_{i=1}^{N} d_i \ln\left(\int_{-\infty}^{\infty} \exp(-\theta(\epsilon)M)g(\epsilon, \sigma_\epsilon^2)d\epsilon\right) \tag{5.30}$$

In the simulator of this expression, when taking the expectation w.r.t. the drawings ϵ_{ir}, we have to condition on $d_i, i = 1, ..., N$.

The problem is that in general there is no rule of drawing random numbers from the conditional density (5.28) without having to evaluate the integral in the denominator of

(5.28), which we do want to avoid by using simulation methods. In practice the only thing one can do is to approximate the procedure of drawing random numbers from (5.28) by using an approximate distribution. The most straightforward choice of approximation is to draw the random numbers from the marginal distribution $\phi(.)$ of u. At the same time this is also the most naive way of approximating the drawing of random numbers from the conditional distribution because any information about the observed values y_i is ignored. We now investigate the bias which is introduced if the moments (5.27) are used with draws u_{ir} from the marginal density $\phi(.)$. First of all note that the bias, i.e. the difference between of the expectation of (5.27) at the true parameter vector and zero, could be avoided by introducing a weight factor $w(u_{ir})$, like in importance sampling, which is the ratio of the true density function (5.28) and the density one actually draws from, in this case the marginal density function $\phi(.)$.

$$w(u_{ir}) = \frac{f(u_{ir}|y_i)}{\phi(u_{ir})} = \frac{f(y_i|u_{ir};\eta)}{\int_{-\infty}^{\infty} f(y_i|\tilde{u};\eta)\phi(\tilde{u})d\tilde{u}} \tag{5.31}$$

which again contains the integral to be simulated. So in order to investigate the bias, we have to look at the consequences of ignoring the weight function. Note, that the expectation of the weight function equals one by construction. As a consequence, the size of the bias is closely related to the variance of the weight function. The smaller the variance of the weight function is, the smaller the bias in the simulated score equations will be, see e.g. Kloek and Van Dijk (1978). (Note that the weight function is identically equal to one if the u_{ir} are drawn from the conditional density function $f(.|y_i)$). The expectation of the square of the weight function is given by:

$$E_u\{(w(u))^2\} = \int_{-\infty}^{\infty} [w(u)]^2\phi(u)du = \int_{-\infty}^{\infty} w(u)f(u|y_i)du \tag{5.32}$$

The addition to the score vector of individuals who accepted a job but whose after spell job characteristics are unobserved can be simulated in exactly the same way. If however, the bounds of the integral are a function of the parameters, like in (5.19), an additional complication arises, i.e. we have to take the derivatives with respect to the bounds. Suppose we have a likelihood addition given by

$$L_i = \ln\left[\int_{a_i(\eta)}^{b_i(\eta)} f(y_i|u;\eta)\phi(u)du\right] \tag{5.33}$$

Then the derivatives are:

$$\frac{\left[\frac{\partial b_i(\eta)}{\partial\eta} f(y_i|b_i(\eta);\eta)\phi(b_i(\eta)) - \frac{\partial a_i(\eta)}{\partial\eta} f(y_i|a_i(\eta);\eta)\phi(a_i(\eta))\right]}{\int_{a_i(\eta)}^{b_i(\eta)} f(y_i|u;\eta)\phi(u)du} + \frac{\int_{a_i(\eta)}^{b_i(\eta)} \partial f(y_i|u;\eta)/\partial\eta\,\phi(u)du}{\int_{a_i(\eta)}^{b_i(\eta)} f(y_i|u;\eta)\phi(u)du} \tag{5.34}$$

The second part can be simulated in exactly the same way as described before, presuming that it is possible to draw random numbers u_{ir} from $(a_i(\eta), b_i(\eta))$, without having to use an acceptance/rejection scheme. The first part represents the derivatives of the probability of being inside the region $(a_i(\eta), b_i(\eta))$, leaving the density unaffected, which can straightforwardly be calculated provided that the bounds are known explicitly. However, this is not the case in the previous section. Therefore, it is proposed to integrate out t, w and h and replace the derivative with respect to the bounds by

$$-\int_{-\infty}^{\infty} \frac{\partial\xi(\epsilon)}{\partial\eta}[1-\exp\{-\theta(\epsilon)M\}]\frac{f(\xi(\epsilon))}{T(\xi(\epsilon))}g(\epsilon,\sigma_\epsilon^2)d\epsilon \tag{5.35}$$

The motivation behind this expression can be found in appendix C. In executing this procedure we have to incur an efficiency loss. However, the expression can still be seen as a good representative for the derivative of the probability of being in the required region. Simulation of the expression is straightforward.

A second simulation estimator can be obtained by simulating the likelihood function rather than looking at the vector of scores. The integrals which appear in (5.25) are replaced by a simulator. Unbiased simulators for the integrals can be obtained by drawing random numbers u_{ir} from the marginal density $\phi(.)$ and calculating

$$\frac{1}{R}\sum_{r=1}^{R} f(y_i|u_{ir};\eta) \tag{5.36}$$

Because of the logarithmic transformation in (5.25) the resulting estimator will be inconsistent for a fixed and small value of R, which actually is the reason why we considered simulation estimators based on the vector of scores before. However, if the simulator is a smooth function of the parameters, which is the case here, the method functions satisfactorily, even for smaller values of R. This is shown in Börsch-Supan and Hajivassiliou (1993) in which they call this method smooth simulated maximum likelihood estimation, where they have added the word "smooth" in order to stress the fact that one needs a simulator which is a smooth function of the parameters rather than a frequency type of simulator in order to get a satisfactory performance.

5.2.3 Specification

In this subsection a specific functional form for the direct utility function is chosen, from which the labour supply function and the indirect utility function can be derived. We use the direct utility function of Hausman (1980).

$$u(y,h;\epsilon) = -\ln(\gamma-\beta h) - \frac{\beta(h-X\delta-\epsilon-\beta y)}{\gamma-\beta h}, \beta<0, \gamma>0 \tag{5.37}$$

where ϵ is assumed to be normally distributed with mean zero and variance σ_ϵ^2 and X is a vector of individual characteristics. Note, that utility is increasing in income as required by assumption 6. It can easily be verified that the second condition of assumption 6 is also satisfied.

Maximizing utility subject to the income equation for the employed yields a linear labour supply function in which the disturbance term equals the unobserved random taste parameter:

$$\bar{h}(w,\mu;\epsilon) = \beta\mu + \gamma w + X\delta + \epsilon \tag{5.38}$$

The virtual wage rate $w_0(\epsilon)$, for which optimal labour supply is exactly equal to zero, defined in (5.8), is given by

$$w_0(\epsilon) = -\frac{\beta\mu + X\delta + \epsilon}{\gamma} \tag{5.39}$$

Note, that there are no positivity constraints on this value. The dynamic reservation wage $\xi(\epsilon)$ will always be higher than this value. Inserting the labour supply function (5.38) into the direct utility function (5.37) yields the expression for the indirect utility function:

$$\nu(w,\mu;\epsilon) = -\ln(\gamma - \beta X\delta - \beta\epsilon - \beta\gamma w - \beta^2\mu) - \beta w \tag{5.40}$$

which is well defined for all $w \geq w_0(\epsilon)$ and therefore for all $w \geq \xi(\epsilon)$.

The wage offer distribution is taken to be log-normal with log-mean $\zeta'x$ and log-variance τ respectively, where x is a vector of individual characteristics. Measurement error v is assumed to be normally distributed with mean zero and variance σ_v^2. The error v allows for difference between optimal labour supply, generated by the labour supply function (5.38), and observed labour supply. The job offer arrival rate can also be made dependent on a vector of characteristics z by specifying

$$\lambda = \exp(\kappa' z) \tag{5.41}$$

Characteristics which may influence the arrival rate of job offers are individual characteristics like age and sector of education, as well as characteristics of the environment of the individual, like the geographical situation.

5.2.4 The data

The data are obtained from the Socio-Economic Panel (SEP), which is a survey, carried out in the Netherlands every half a year in April and October by the Central Bureau of Statistics (CBS). In the survey, the participating individuals are asked to report their occupational status for every month of the past six months. The data which are used are those of the wave of the survey in October 1985 up to the wave in October 1987, which implies that the observation period is two and a half years. Selected, are male individuals who reported to be unemployed in any month during the observation period, which means that the sample partly has a stock character and partly a flow character. It has been determined how many months they remained unemployed, and for the individuals in the stock sample the backward recurrence times are also gathered. For most of the individuals whose unemployment spell ended during the observation period, data on their after spell hourly wage rate and the weekly number of hours are available. The sample consists of 516 individuals. The number of complete unemployment spells is 272. The remaining 244 spells are truncated. For 211 of the 272 individuals whose unemployment spells are completed the after spell job characteristics, i.e. the hourly wage rate and the weekly number of working hours, are observed.

Looking at the data on the completed unemployment spells, we see that peaks in the frequencies of unemployment durations can be found every six months in the months where the survey on the preceding six months year is conducted. In figure 5.1 the frequencies of the completed unemployment spells of the flow sample and the forward recurrence times of the stock sample are plotted and the peaks are at duration levels of 6, 12, 18 and 24 months. The peak observations are apparently the result of misreporting and as a result the data on unemployment durations in units of one month are unreliable. Therefore, the durations in units of one month are divided into groups of six months. The likelihood contribution then becomes the probability that the unemployment spell ends somewhere in the observed interval of six months.

In table 5.1 some sample statistics are given. In the survey there are five levels of education where level 1 is the lowest and level 5 the highest. The mode of the level of education is 2. We have divided the Netherlands into four regions. Region 1 is the most strongly industrialized part of the Netherlands which includes the larger cities. Region 4 is the least industrialized part of the Netherlands with a relatively low population density and a sizeable agricultural sector. Region 3 is the south of the Netherlands which contains some large companies and agricultural industry. In region 2 (the east) there is a mix of industry and agriculture. Apart from having information about the level of education we have information available about the type or sector of education. Sector 1 is the technical sector which includes chemistry, physics, mathematics and biology, sector 2 includes the economic and administrative directions, sector 3 is general education and sector 4 includes services.

In table 5.2, estimates obtained with simulated maximum likelihood, using $R = 10$ drawings are presented. According to the estimates of the utility parameters, optimal labour supply is not very sensitive with respect to non-labour income and the wage rate. The wage elasticity of labour supply is 0.0996. Family size has a positive effect on the optimal amount of labour supply. Age has a negative effect on the job offer arrival rate, i.e. the older one is, the fewer job offers can be expected. The nationality dummy, which is one for those who do not have the Dutch nationality, is negative, indicating that having the Dutch nationality influences the job offer arrival rate positively. The regional dummies have a positive sign, which means a higher job offer arrival rate for people living outside the agricultural region 4. Only the dummies for region 1 and region 3 are significant. Living in the industrial west is associated with a higher job offer arrival rate than living in the rest of the country. Only the sectoral dummy for sector 3 which includes those individuals who only had a general type of education, i.e. no specialisation in a certain profession, is significant. The highest wages are offered to the individuals with the highest level of education.

The parameter estimate of ρ is 0.00492. As unemployment duration is measured in months, this means that the monthly discount rate is 0.49%, or equivalently, the discount rate is 5.9% per year.

As noted before, estimates obtained by Narendranathan and Nickell (1985) and Van den Berg (1990c) showed that unemployment, i.e. total leisure, was assigned a lower utility value than employment. Their reservation wage equation is of the form

$$u(\xi + \mu) - \omega u(b + \mu) = \frac{\lambda}{\rho} \int_{\xi}^{\infty} (u(x) - u(\xi))\, dF(x) \tag{5.42}$$

in which ω is the utility parameter whose estimates have been found to be less than one. Now note that the right hand side of (5.42) is always positive from which we can derive that

$$u(\xi + \mu) > \omega u(b + \mu) \tag{5.43}$$

Now $\omega \geq 1$ implies that the reservation wage rate ξ is always higher than the benefit rate b. However, if $\omega < 1$, the reservation wage rate is allowed to be lower than the benefit rate. Of course, a value of the reservation wage rate which is actually below the benefit rate indicates that unemployment must have a lower utility value than employment. Our specification has enough flexibility for the labour income at the reservation wage rate to be below the income from benefits, but at the same time whether unemployment is valued lower than employment cannot be checked directly by looking at a single parameter. Therefore, it makes sense to run a simulation by drawing random preferences ϵ for each individual, calculating the reservation wage rate $\xi(\epsilon)$, computing the optimal labour income at the reservation wage rate if $\xi(\epsilon) > 0$, i.e. $\xi(\epsilon)\bar{h}(\xi(\epsilon), \mu; \epsilon)$ and comparing it with the benefit income b. This procedure was repeated 100 times and the result is that for 87% of the sample the reservation income is lower than the benefit income. The reason for this high percentage is the fact that 70% of the individuals in the sample is working about 40 hours a week. In the context of this model, in which individuals are assumed to choose their amount of labour supply optimally, this means that these individuals are working 40 hours a week because they like to work 40 hours a week. As a consequence, the virtual wage rate $w_0(\epsilon)$, given in (5.39), is relatively low. Although the reservation wage rate $\xi(\epsilon)$ always exceeds the virtual wage rate, the reservation wage rate is close to the virtual wage rate as the estimate of β is close to zero. This result is consistent with the findings of Narendranathan and Nickell (1985) and Van den Berg (1990c) who have estimated a value of ω in (5.42) that is smaller than 1, i.e. unemployment lowers utility.

Table 5.1 Sample statistics

variable	mean	standard deviation
working hours (hours/week)	39.0	9.1
after tax hourly wage rate (guilders/hour)	10.1	4.5
benefits (guilders/week)	289.6	108.2
non-labour income (guilders/week)	80.4	188.8
age	31.2	11.8
family size (persons)	3.2	1.7
education level	mode 2	
Dutch nationality	93.6%	
region 1 (industrialized west)	31.6%	
region 2 (east)	29.3%	
region 3 (south)	26.6%	
region 4 (agricultural)	12.6%	
sector of education 1 (technical)	30.4%	
sector of education 2 (economic/administrative)	8.7%	
sector of education 3 (no specialization)	48.8%	
sector of education 4 (services)	12.0%	

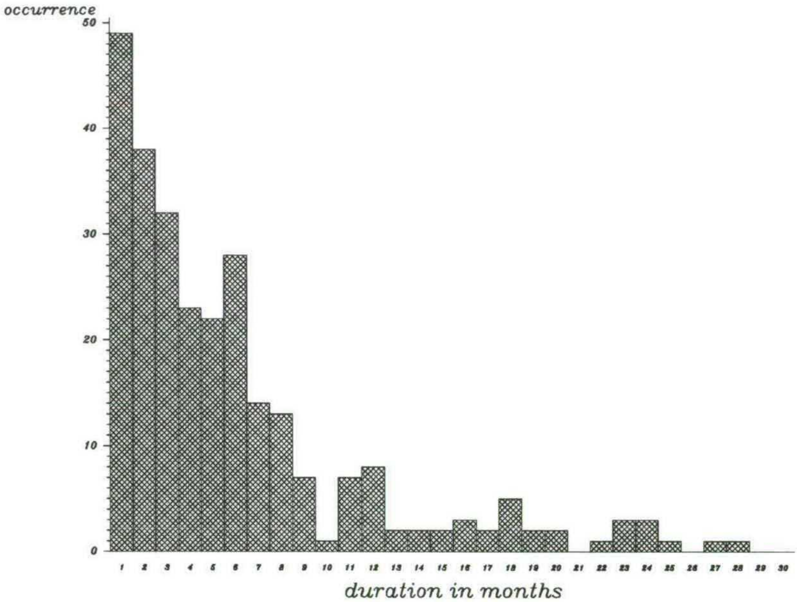

Figure 5.1: Unemployment duration in months, completed spells and forward recurrence times

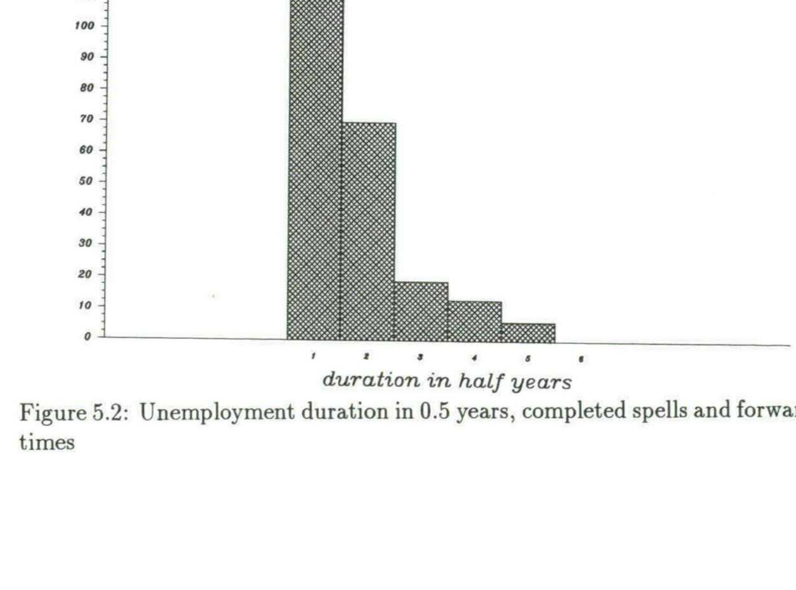

Figure 5.2: Unemployment duration in 0.5 years, completed spells and forward recurrence times

TABLE 5.2 ESTIMATION RESULTS WITH SIMULATED MAXIMUM LIKELIHOOD, R = 10

THE UTILITY PARAMETERS		
β	-2.479×10^{-5}	1.301×10^{-5}
γ	0.378	0.251
δ_1, Constant	32.076	2.454
δ_2, Log family size	2.064	1.790
σ_ϵ	1.488	0.0286
THE JOB OFFER ARRIVAL RATE PARAMETERS		
κ_1, Constant	0.462	0.320
κ_2, Log age	−0.215	0.478
κ_3, Nationality	−0.958	0.351
κ_4, Region 1	0.676	0.202
κ_5, Region 2	0.312	0.195
κ_6, Region 3	0.470	0.204
κ_7, Sector 1	−0.0748	0.218
κ_8, Sector 2	−0.293	0.304
κ_9, Sector 3	−0.546	0.214
κ_{10}, Square of log age	−0.235	0.604
RATE OF TIME PREFERENCE		
ρ	0.00492	1.008
THE WAGE DISTRIBUTION		
ζ_1, Constant	−12.084	3.135
ζ_2, Log age	8.088	1.874
ζ_3, Square of log age	−1.104	0.279
ζ_4, educ1	−0.329	0.0856
ζ_5, educ2	−0.289	0.0789
ζ_6, educ3	−0.165	0.0797
τ	0.293	0.0104
DISTR. OF MEASUREMENT ERROR		
σ_v	0.312	0.0909

5.3 Job search with hours restrictions

5.3.1 The model

In general when looking at pictures of frequency distributions of labour supply data one can hardly maintain the assumption that these data can be described by a standard neo-classical microeconometric labour supply model. Spikes can be observed in the empirical probability mass functions at weekly numbers of hours, like 40. In the literature there are different explanations for the presence of this concentration of labour supply at certain amounts of hours, but what all these explanations have in common is that they, in one way or another, refer to the demand side of the labour market. The theory of compensating wage differentials, see for example Abowd and Ashenfelter (1981), assumes that employers offer hours quantities, like 40, and that employers are compensated by means of a higher wage rate for the loss in utility they experience by being over- or underemployed. Because of this compensation the individual is not constrained in the sense that the same utility level is acquired as in the case in which the individual could determine hours by utility maximization without wage compensation. The emphasis is on the wage function, which describes wages as a function of hours, rather than on a labour supply function. Although it is plausible that individuals within their job are compensated for working overtime hours, over and above the amount of hours agreed to in the labour contract, it is doubtful whether wage rates at spike levels are systematically higher. Especially in an economy with unemployment, the employers do not have to raise wages in order to satisfy their demand for labour.

The second point of view, then, is that individuals are not compensated for being under- or overemployed. Moffitt (1982) recognizes that relatively few jobs exist with hours in the part time range. He models this by adapting the standard Tobit model. Dickens and Lundberg (1985) assume that individuals are constrained by the hours that are offered to them by the employers. Their framework is static and assumes that at a point in time the individuals receive a random amount of job offers (possibly zero) from which they choose the offer which yields the highest level of utility. This utility level is compared with the utility of not working, after which it is decided whether or not to accept the job. All job offers have the same gross wage rate but may differ in the number of hours. The hours offer distribution is modelled by means of a discrete probability distribution. The idea was applied by Tummers and Woittiez (1991), Van Soest, Woitiez and Kapteyn (1990). In the previous chapter the assumption of having the same wage rate for every job offer has been dropped. In fact, the model in chapter 4 can be interpreted as a static version of the model presented here. Blau (1991) presents a dynamic search model in which a job offer has both a wage and an hours component. In his paper a comparison is made between the expected utility maximizing model, in which both wages and hours play a role, and the expected wealth maximizing model, in which the individual decision maker is concerned about wage income only. The choice of a separable utility function enabled Blau (1991) to nest the latter model into the first, and evidence in favour of the more general utility maximizing model was found. However, in the previous section it has been shown that one of the underlying assumptions of the wealth maximizing model is the unrestricted choice of labour, whereas in this section it is shown that the utility maximizing model arises due to hours restrictions. As a consquence, the two models can never be nested once the assumption on labour supply is allowed for explicitly.

Instead of jobs arriving at a given point in time, as in the static models of labour supply with hours restrictions, we now assume that job offers arrive sequentially, like in the previous section. But in contrast, a job offer now consists of both a wage rate and an amount of weekly working hours. Individuals will evaluate the utility level of jobs, comparing them with the value of not accepting the job taking into account possible future job offers. Individuals are constrained in hours offered to them and therefore, unlike the standard job search framework, the wage rate is not the only job characteristic which is needed to decide whether or not to accept a job. Instead of having a reservation wage it can be derived that there exists a unique reservation utility level. Jobs, i.e. wage-hours combinations, which generate a utility level which is higher than the reservation utility level are accepted, those with a utility level below are rejected.

More formally, assumption 1 in section 5.2.1, which states that individuals are utility maximizers in the neo-classical sense, is replaced by the assumption that hours arrive from an hours offer distribution, just like the wage rate. The equation for the reservation utility level is derived in appendix B. It is given by

$$\bar{u}(\epsilon) = u(b+\mu,0;\epsilon) + \frac{\lambda}{\rho}\int_0^\infty \int_{\xi(h,\bar{u}(\epsilon);\epsilon)}^\infty [u(wh+\mu,h;\epsilon) - \bar{u}(\epsilon)] f(w,h)dwdh \tag{5.44}$$

where $\bar{u}(\epsilon)$ is the reservation utility level at random taste parameter ϵ, $f(w,h)$ is the joint wage-hours offer distribution and $\xi(h,\bar{u}(\epsilon);\epsilon)$ is the wage rate which generates utility level $\bar{u}(\epsilon)$ at hours h. The wage rate $\xi(h,\bar{u}(\epsilon);\epsilon)$ can be interpreted as an hours dependent reservation wage rate. The sign of the derivative of $\xi(h,\bar{u}(\epsilon);\epsilon)$ with respect to hours h depends on whether the individual is over- or underemployed at wage level $\xi(h,\bar{u}(\epsilon);\epsilon)$ and the given value of hours. If an individual is overemployed, an increase in hours moves utility further away from its unconstrained value and consequently the hours dependent reservation wage rate rises, thereby increasing the reluctancy to accept a job. And vice versa for underemployment. As a result the reservation wage rate does not depend on hours if the individual is not constrained in the choice of hours, which leads back to the model in section 5.2.

In case of a discrete hours offer distribution the outer integral is replaced by summation.

The resulting expression for the escape rate becomes

$$\theta(\epsilon) = \lambda \int_0^\infty \int_{\xi(h,\bar{u}(\epsilon);\epsilon)}^\infty f(w,h)dwdh \tag{5.45}$$

Taking into account the different functional form of the escape rate, the likelihood contributions of individuals who did not accept a job during the observation period and those who did, but whose after spell wage rate and hours are not observed, are the same as in the previous section in (5.12) and (5.21) respectively. For the individuals whose after spell wage-hours pair is observed to be (w,h), and, again, the observed wage rate is truncated, conditional on hours and on ϵ, which leads to the likelihood contribution, conditional on ϵ:

$$\begin{aligned} l_{co}(\eta|\epsilon) &= \theta(\epsilon)\exp\{-\theta(\epsilon)t\}f(w,h)/Q(\epsilon), \\ & 0 < t < M, h > 0, w > \max\{0, \xi(h, \bar{u}(\epsilon); \epsilon)\} \end{aligned} \tag{5.46}$$

$$= 0 \text{ otherwise} \tag{5.47}$$

where $Q(\epsilon)$ is the truncation probability, given by

$$Q(\epsilon) = \int_0^\infty \int_{\xi(h,\bar{u}(\epsilon);\epsilon)}^\infty f(w,h)dwdh \tag{5.48}$$

The unconditional likelihood function is obtained by integrating over those values of ϵ for which the utility level evaluated in (w, h) exceeds the reservation utility level.

$$l_{co}(\eta) = \int_{I_{wh}} \theta(\epsilon)\exp\{-\theta(\epsilon)t\}\frac{f(w,h)}{Q(\epsilon)}g(\epsilon,\sigma_\epsilon^2)d\epsilon, t > 0, h > 0, w > 0 \tag{5.49}$$

where the set I_{wh} is given by

$$I_{wh} = \{\epsilon | u(wh + \mu, h|\epsilon) > \bar{u}(\epsilon)\} \tag{5.50}$$

The set $I(w, h)$ has to be non-empty for every (w, h) with $w > 0, h > 0$.

For the individuals in the stock sample the expression (5.22) for the probability of t exceeding p can be used, taking into account that $\theta(\epsilon)$ is now determined by (5.45).

5.3.2 Specification

For the utility function we take the specification of section 5.2.3 given in (5.34). The spikes in the empirical hours distribution suggest the use of a discrete hours offer distribution of the Dickens and Lundberg (1985) type. However, we are hampered by the fact that 68% of the observed labour supply is at values of 38 or 40 hours a week, making it impossible to identify probabilities of low or high levels of labour supply. Therefore, only a rough distinction is made between part time jobs (32 hours or less), full time jobs of normal levels (33 to 44 hours a week) and full time jobs of high levels (more than 44). The hours are categorized in classes of four, which yields 20 classes of hours from 1 to 80. The probability of getting a job offer from class l is given by

$$P(h = h_l) = p_l, l = 1, ..., m, m = 20 \tag{5.51}$$

The probabilities can parametrized by

$$p_l = \frac{\mu_l}{\sum_{j=1}^m \mu_j}, l = 1, ..., m, m = 20 \tag{5.52}$$

where the normalization

$$\mu_m = 1 \tag{5.53}$$

is made. The distinction made between part time and full time jobs can be expressed in the following restrictions:

$$\mu_1 = ... = \mu_8 \quad : \text{part time jobs} \tag{5.54}$$
$$\mu_9 = \mu_{10} = \mu_{11} \quad : \text{normal full time jobs} \tag{5.55}$$
$$\mu_{12} = ... = \mu_{20} \quad : \text{high level jobs} \tag{5.56}$$

Again, it is assumed that wages arive from a log-normal wage offer distribution with log-variance τ and log-mean $\zeta' x$. The wage offer density function is

$$f(w) = \frac{1}{\sqrt{2\pi}\tau}\frac{1}{w}\exp\left\{-\frac{1}{2\tau^2}[\ln w - \zeta' x]^2\right\}, 0 < w < \infty \tag{5.57}$$

Using the discrete hours offer distribution and the wage offer distribution the reservation utility equation becomes

$$\bar{u}(\epsilon) = u(b+\mu, 0; \epsilon) + \frac{\lambda}{\rho}\sum_{l=1}^{m} p_l \int_{\xi(h_l,\bar{u}(\epsilon);\epsilon)}^{\infty} [u(wh_l + \mu, h_l; \epsilon) - \bar{u}(\epsilon)] f(w) dw \tag{5.58}$$

A similar expression can be found for the escape rate $\theta(\epsilon)$:

$$\theta(\epsilon) = \lambda \sum_{l=1}^{m} p_l \int_{\xi(h_l,\bar{u}(\epsilon);\epsilon)}^{\infty} f(w) dw = \lambda \sum_{l=1}^{m} p_l \bar{F}(\xi(h_l, \bar{u}(\epsilon); \epsilon)) \tag{5.59}$$

In the previous section, the assumption has been made that only preferences depend on unobserved individual characteristics ϵ. Now we will assume that the job offer arrival rate also depends on an unobserved random variable q, known to the individual, which is independent of ϵ and which is normally distributed with mean zero and variance σ_λ^2. Inclusion of q in the job offer arrival rate at the same time ensures that the set I_{wh} defined in (5.50) is non-empty, i.e. for every observed job offer (w, h) there exists an ϵ such that $u(wh + \mu, h; \epsilon) > \bar{u}(\epsilon)$ which is consistent with the fact that we do observe (w, h). The job offer arrival rate becomes

$$\lambda(q) = \exp(\kappa' z + q) \tag{5.60}$$

In table 5.3, estimates obtained with simulated maximum likelihood (SSML) using $R = 10$ drawings are presented. This model generates estimates of β and γ which are clearly different from the estimates of the model in the previous section. The reason for this is that in the neo-classical model in section 5.2 the parameter estimates of the utility function are mainly determined by the labour supply data, because of the assumption that hours are chosen by the individuals according to their preferences. In the neo-classical model the presence of a spike at 40 hours a week in the labour supply data can only be explained by inelastic labour supply, whereas in the present model with hours restrictions the alternative explanation for the spike is given by the presence of demand side restrictions.

Again we see a positive effect of family size. The variance σ_λ of the unobserved heterogeneity in the job offer arrival rate is not very large. From the regional dummies, only the dummy variable for region 1, the industrialized western part of the Netherlands, is significant. The only significant sectoral dummy is that for sector 3, which is the sector of individuals who are not specialized in a certain profession. The sign of the estimate is negative.

The estimate of ρ is 0.688 which is equivalent to a monthly discount rate of almost 70%, which is rather high. In section 5.2.1, however, it was recognized that one should be careful in interpreting the estimate of ρ. For example, if assumption 5 in section 5.2.1 is replaced by the assumption that there is an exogenous layoff rate, the estimate of ρ should be interpreted as the sum of this layoff rate and the discount rate. Equivalently, the ignorance of other possibly relevant labour market states changes the interpretation of the estimate of ρ in a similar way, see e.g. Van den Berg (1990c).

To compare the labour income evaluated at the reservation wage rate, we run the same kind of simulation procedure as in the previous section. We compare the expected reservation income with the benefit level b. The expected reservation income has been defined by:

$$\sum_{l=1}^{m} p_l \xi(h_l, \bar{u}(\epsilon); \epsilon) h_l l(\xi(h_l, \bar{u}(\epsilon); \epsilon) > 0) \tag{5.61}$$

in which $l(.)$ is the indicator function. The number of replications is 100. For 98.7% of the number of individuals in the sample, the expected reservation income in the sample is higher than the benefit level. This result is not consistent with a disutility of unemployment ($\omega < 1$ in (5.42)). An explanation for the low reservation income in the neo-classical model has already been given in section 5.2.4. Apart from that explanation, there is an additional explanation for why the reservation income defined in (5.61) may be higher than the reservation income in section 5.2.4. In section 5.3.1 it has been explained that the larger the deviation of offered hours from optimal labour supply is, the higher the hours dependent reservation wage rate $\xi(h, \bar{u}(\epsilon); \epsilon)$ will be. The hours dependent reservation wages are higher because the individual gets hours offers that do not coincide with optimal labour supply. So, even if the reservation wages were evaluated in the same parameter values for both models, the reservation wage of the model with hours restrictions will be higher than the reservation wage of the neo-classical model. As the hours offer probabilities are large in the full time range, the reservation income in (5.61) is likely to be higher than the neo-classical reservation income as well.

TABLE 5.3 ESTIMATION RESULTS WITH SIMULATED MAXIMUM LIKELIHOOD, R = 10

THE UTILITY PARAMETERS		
β	−0.0123	0.0107
γ	1.726	0.968
δ_1, Constant	27.478	10.523
δ_2, Log family size	8.576	4.386
σ_ϵ	21.334	14.945
THE JOB OFFER ARRIVAL RATE PARAMETERS		
σ_λ	0.000963	0.228
κ_1, Constant	1.239	0.818
κ_2, Log age	−0.327	0.585
κ_3, Nationality	−0.872	0.378
κ_4, Region 1	0.720	0.257
κ_5, Region 2	0.182	0.247
κ_6, Region 3	0.232	0.256
κ_7, Sector 1	−0.00988	0.250
κ_8, Sector 2	−0.574	0.342
κ_9, Sector 3	−0.572	0.248
κ_{10}, Square of log age	−0.230	0.114
RATE OF TIME PREFERENCE		
ρ	0.688	3.576
THE WAGE DISTRIBUTION		
ζ_1, Constant	−12.502	3.181
ζ_2, Log age	8.287	1.903
ζ_3, Square of log age	−1.129	0.284
ζ_4, educ1	−0.320	0.0823
ζ_5, educ2	−0.276	0.0776
ζ_6, educ3	−0.144	0.0111
τ	0.288	0.0111
PROBABILITIES OF HOURS DISTRIBUTION		
$p_1 = ... = p_8$ part time	0.0204	0.00373
$p_9 = ... = p_{11}$ full time	0.251	0.0114
$p_{12} = ... = p_{20}$ high level jobs	0.00944	0.00217

5.4 Residual analysis and simulated frequencies

In order to be able to say something about the performance of the two models we will do some residual analysis. An overview of residual analysis of duration models can be found in Lancaster (1990) and Cox and Oakes (1984). There are two points according to which the standard residual analysis in duration models cannot be applied directly to our models. The first point is that this analysis is usually performed in the context of a flow sample, whereas we have a sample which consists of a flow subsample as well as a stock subsample. Therefore, in the analysis of the residuals we restrict ourselves to the individuals who are in the flow subsample. The second point is that residual analysis is usually performed on models which do not contain a random unobserved heterogeneity component. This problem is solved in the following way. The marginal density of observed duration t is given by $f(t)$.

$$f(t) = \int_{-\infty}^{\infty} \theta(\tilde{\epsilon}) \exp\{-\theta(\tilde{\epsilon})t\} g(\tilde{\epsilon}, \sigma_\epsilon^2) d\tilde{\epsilon}, 0 < t < \infty \tag{5.62}$$

The density of the unobserved heterogeneity variable ϵ, conditional on observed duration t is given by

$$g(\epsilon|t) = \frac{\theta(\epsilon) \exp\{-\theta(\epsilon)t\} g(\epsilon, \sigma_\epsilon^2)}{\int_{-\infty}^{\infty} \theta(\tilde{\epsilon}) \exp\{-\theta(\tilde{\epsilon})t\} g(\tilde{\epsilon}, \sigma_\epsilon^2) d\tilde{\epsilon}}, -\infty < \epsilon < \infty \tag{5.63}$$

Now draw a random number ϵ_r from $g(.|t)$. This can be done by using the inversion method, see e.g. Devroye (1986). Then the pair (t, ϵ_r) can be seen as a joint draw from

$$f(t)g(\epsilon|t) = \theta(\epsilon) \exp\{-\theta(\epsilon)t\} g(\epsilon, \sigma_\epsilon^2), 0 < t < \infty, -\infty < \epsilon < \infty \tag{5.64}$$

Consider the following transformation:

$$\begin{aligned} \vartheta &= \theta(\epsilon)t \\ \epsilon &= \epsilon \end{aligned} \tag{5.65}$$

The jacobian of the transformation is $1/\theta(\epsilon)$. The joint density of ϑ and ϵ becomes

$$f\left(\frac{\vartheta}{\theta(\epsilon)}\right) g\left(\epsilon|\frac{\vartheta}{\theta(\epsilon)}\right) = \exp\{-\vartheta\} g(\epsilon, \sigma_\epsilon^2), 0 < \vartheta < \infty, -\infty < \epsilon < \infty \tag{5.66}$$

which implies that ϑ is exponentially distributed with parameter 1, which enables us to apply the standard residual analysis. Summarizing, the procedure is:

- Draw ϵ_r from $g(.|t)$, where $g(.|t)$ is evaluated in the parameter estimates and t is observed duration
- Calculate $\hat{\vartheta}_r = \hat{\theta}(\epsilon_r)t = \theta(\epsilon_r, \hat{\eta})t$ in which $\hat{\theta}(\epsilon_r)$ is the hazard rate evaluated in the parameter estimate $\hat{\eta}$. $\hat{\vartheta}_r$ is the simulated residual. For every individual, several residuals can be calculated by drawing several random numbers.
- Calculate the Kaplan-Meier estimate of the survivor function in the residuals $\hat{\vartheta}_{ir}, i = 1, ..., N, r = 1, ..., R$, in which the subindex i is over individuals and the subindex r is over different draws from the conditional density.

The residuals can be plotted against minus the logarithm of the Kaplan-Meier estimate. If the parametric model is correctly specified, the plot would be approximately a 45 degree line. Various forms of misspecification of the hazard rate can cause the residual plot to deviate from the 45 degree line. Lancaster (1990) shows that omitting an unobserved heterogeneity factor in the hazard rate leads to underdispersion, i.e. the plot will be below the 45 degree line. Particularly interesting for our application are the deviations which are caused by wrongly assuming that the hazard rate is constant, i.e. the stationarity assumption. Ridder (1987) shows that when the data inhibit positive duration dependence, whereas a constant hazard model is estimated, the residual plot will be above the 45 degree line. In case of negative duration dependence of the hazard rate, the reverse holds.

For every individual, five residuals have been simulated. Figure 5.3 shows the plot of minus the logarithm of the Kaplan-Meier estimate versus the residuals for model 1, whose estimates were presented in table 5.2. Figure 5.4 shows the same plot for model 2, based on the estimates in table 5.3. For both models the plot is above the 45 degree line, indicating that there could be positive duration dependence. However, for model 2 the deviations are much less severe than for model 1. In the figures 5.5 and 5.6 the Kaplan-Meier survivor functions and the exponential survivor functions are plotted. The plots reveal that even though the parameter estimates of the job offer arrival rate show similar effects for both models, the misspecification of model 2, in which individuals are faced by hours restrictions on the labour market, is much less severe than in model 1.

Figure 5.7 shows the frequency distribution of simulated hours, conditional on observed wages, for model 1, the model without hours restrictions, and the sample frequencies of observed hours. In this figure, hours are divided in three categories, i.e. part time jobs, full time jobs and high level jobs, and the frequencies shown in the figure are the frequencies of these categories. The distinction between the categories has been described in section 5.3.2. The density of hours, conditional on wages can be derived from (5.19). It is given by

$$r(h|w) = \frac{\int_{I_w} r(h|w,\epsilon)\frac{1}{T(\epsilon)}g(\epsilon,\sigma_\epsilon^2)d\epsilon}{\int_{I_w}\frac{1}{T(\epsilon)}g(\epsilon,\sigma_\epsilon^2)d\epsilon}, 0 < h < \infty \tag{5.67}$$

To simulate labour supply, conditional on the observed wage rate, ϵ is drawn from $g(\epsilon,\sigma_\epsilon^2)$ restricted to the region I_w. Optimal labour supply can then be calculated from (5.13) or, more specific, from (5.38), after which a measurement error v is drawn to simulate observed labour supply by (5.14).

Figure 5.8 shows the frequency distribution of simulated hours for model 2, the model with hours restrictions, and the sample distribution of observed hours. The distribution of hours for the second model can be derived from (5.49). It is given by

$$p_l \int_{-\infty}^{\infty} \frac{\bar{F}(\xi(h_l,\bar{u}(\epsilon);\epsilon))}{Q(\epsilon)} g(\epsilon,\sigma_\epsilon^2)d\epsilon, l = 1,...,m \tag{5.68}$$

These are the probabilities of observing $h_l, l = 1,...,m$. A value for ϵ is drawn from $g(\epsilon,\sigma_\epsilon^2)$, which makes it possible to calculate $\bar{u}(\epsilon)$. Then

$$\frac{\bar{F}(\xi(h_l,\bar{u}(\epsilon);\epsilon))}{Q(\epsilon)} p_l \tag{5.69}$$

is calculated for $l = 1, ..., m$. Uniform random variates on the interval $(0,1)$, together with the probabilities in (5.69), are used to determine a simulated value for h_l. Comparing figure 5.7 with figure 5.8, we see that for the model with hours restrictions the sample distribution of observed hours is fitted much better than for the model without hours restrictions.

Figure 5.9 shows the frequencies of simulated wages and the sample frequencies of wages for model 1. Again, the density of observed wages for model 1 can be determined from (5.19). It is given by

$$\int_{I_w} \frac{f(w)}{T(\epsilon)} g(\epsilon, \sigma_\epsilon^2) d\epsilon, 0 < w < \infty \tag{5.70}$$

A wage rate is simulated by first drawing a value for ϵ from $g(\epsilon, \sigma_\epsilon^2)$ and calculating $\xi(\epsilon)$, after which a wage rate is drawn from $f(w)$, restricted to the region $(\xi(\epsilon), \infty)$. For model 2 the frequencies of simulated wages can be found in figure 5.10. The marginal distribution of wages, derived from (5.49) is

$$\sum_{l=1}^{m} p_l \int_{I_{wh_l}} \frac{f(w)}{Q(\epsilon)} g(\epsilon, \sigma_\epsilon^2) d\epsilon, 0 < w < \infty \tag{5.71}$$

A value for h_l is drawn from the hours offer distribution, ϵ is drawn from $g(\epsilon, \sigma_\epsilon^2)$ after which a wage rate is drawn from $f(w)$ restricted to the region $(\xi(h_l, \bar{u}(\epsilon); \epsilon), \infty)$. Comparing the figures 5.9 and 5.10, we see that there is not much difference between the two models.

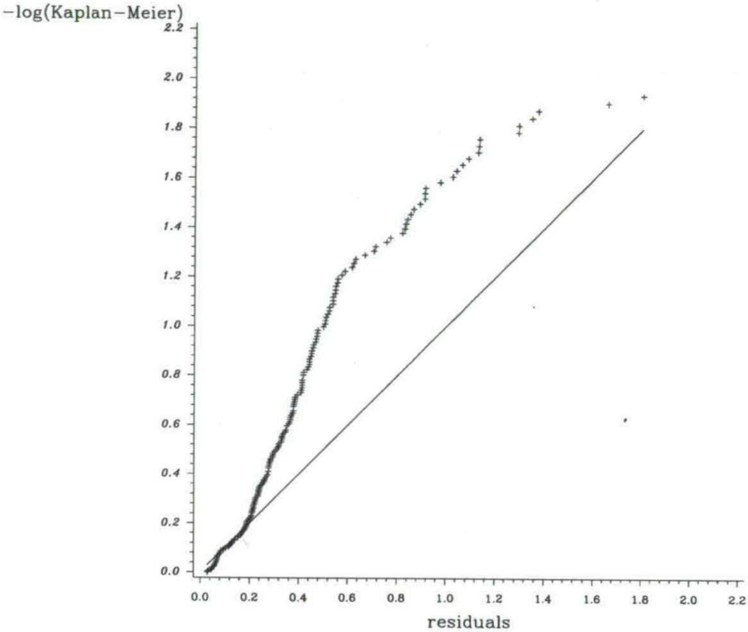

Figure 5.3: Model 1: -log(Kaplan-Meier) vs. residuals

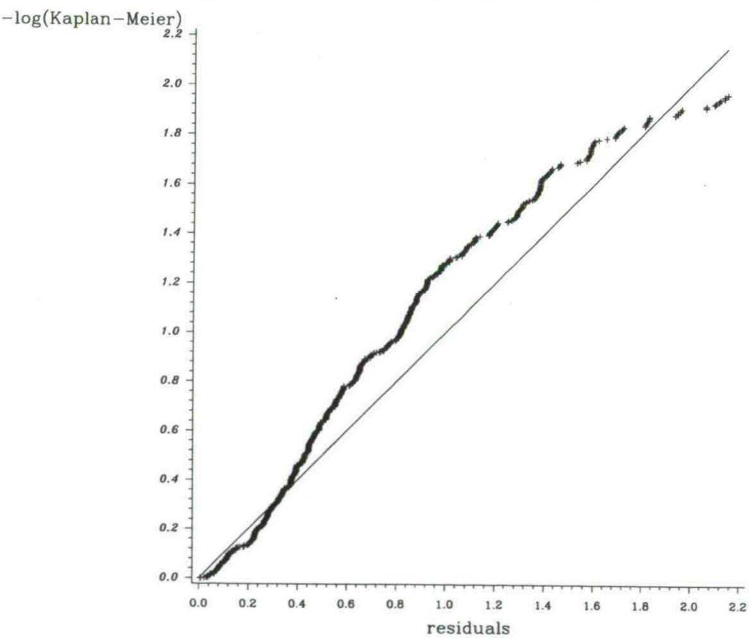

Figure 5.4: Model 2: -log(Kaplan-Meier) vs. residuals

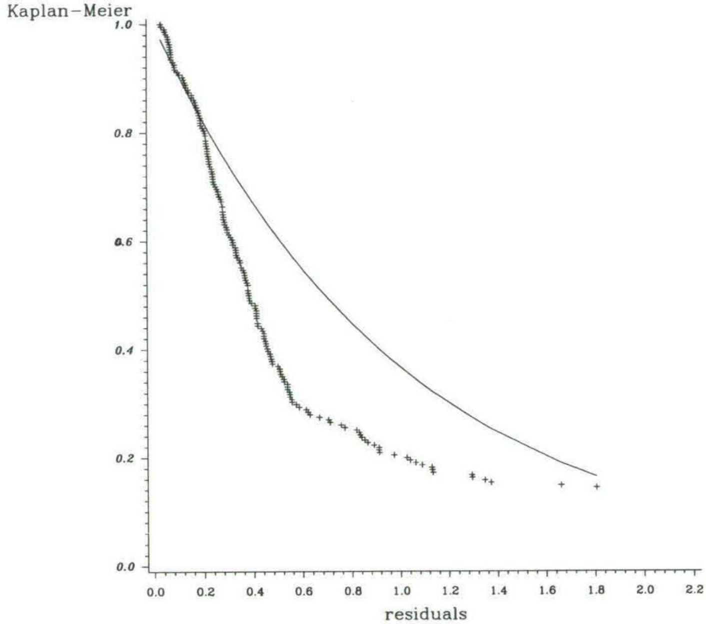

Figure 5.5: Model 1: Kaplan-Meier survivor function

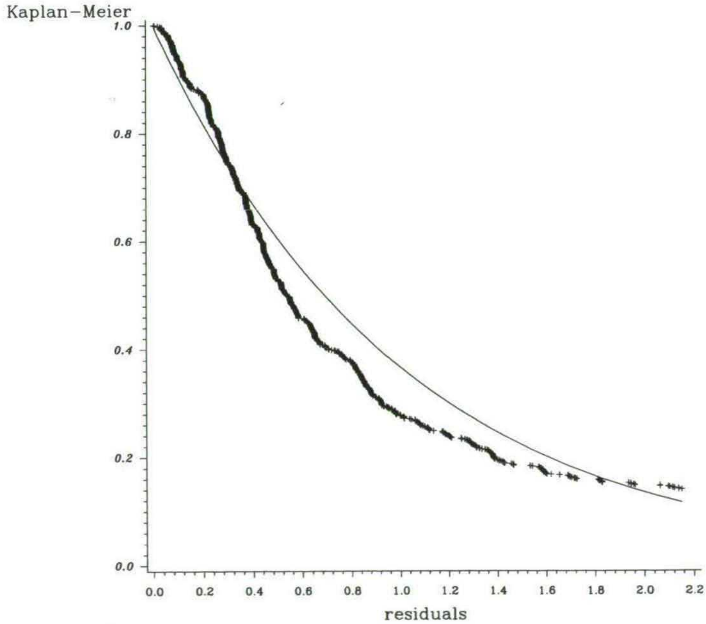

Figure 5.6: Model 2: Kaplan-Meier survivor function

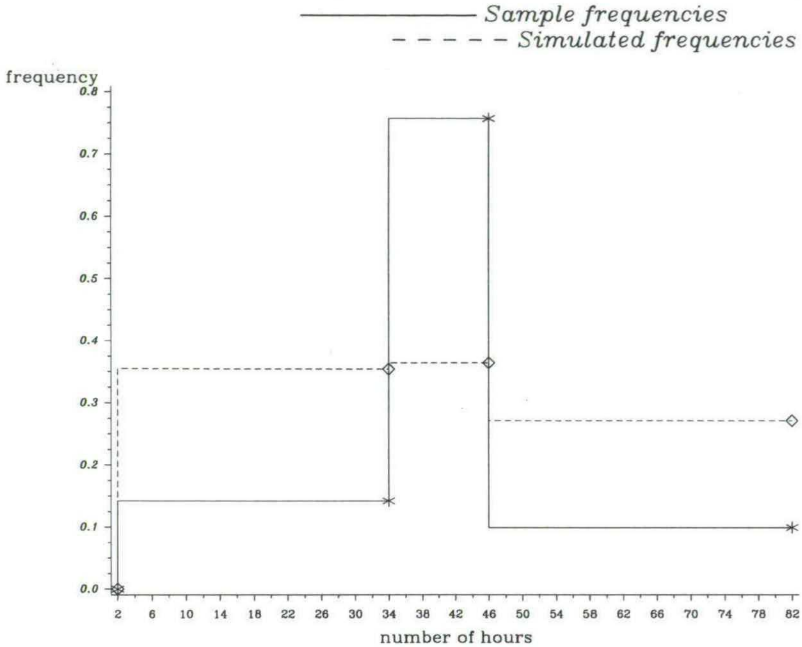

Figure 5.7: Model 1: Frequencies of working hours per week

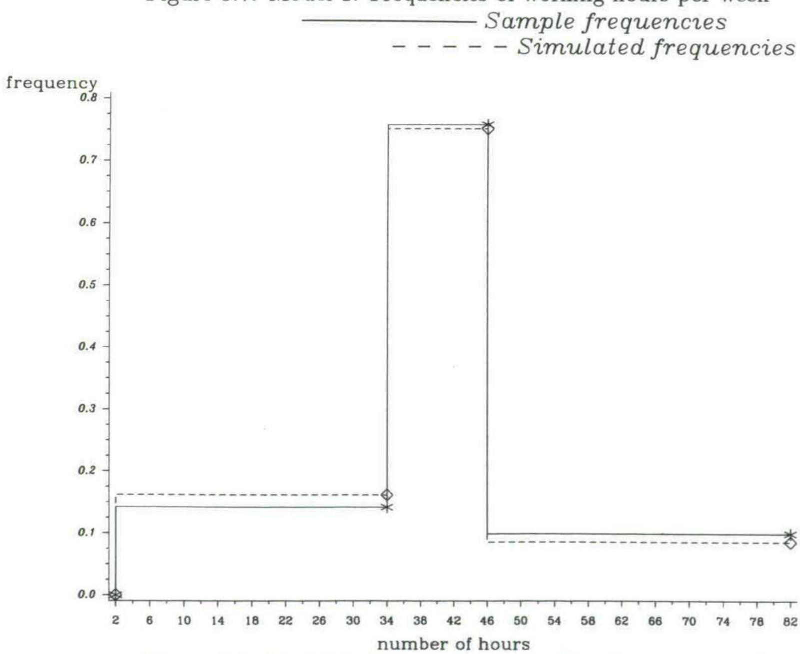

Figure 5.8: Model 2: Frequencies of working hours per week

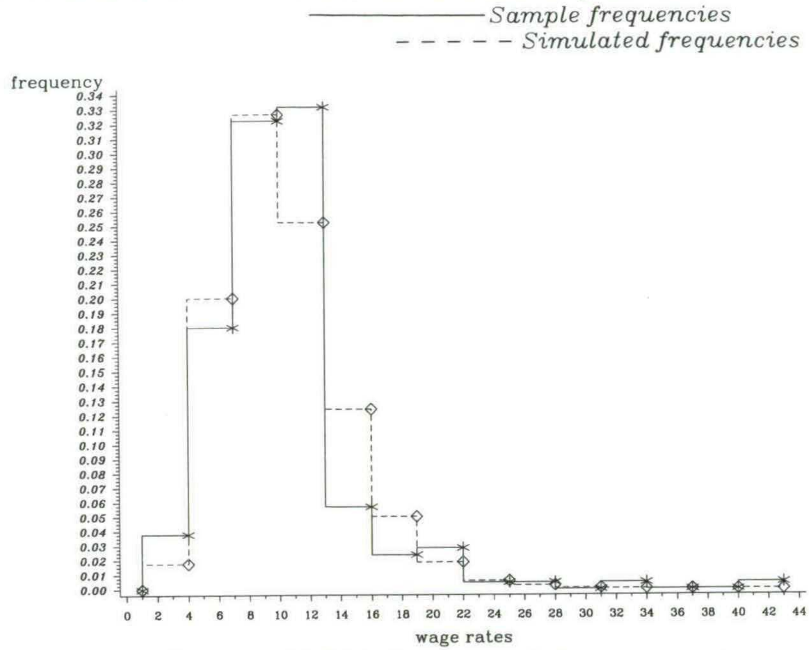

Figure 5.9: Model 1: Frequencies of after tax wage rates

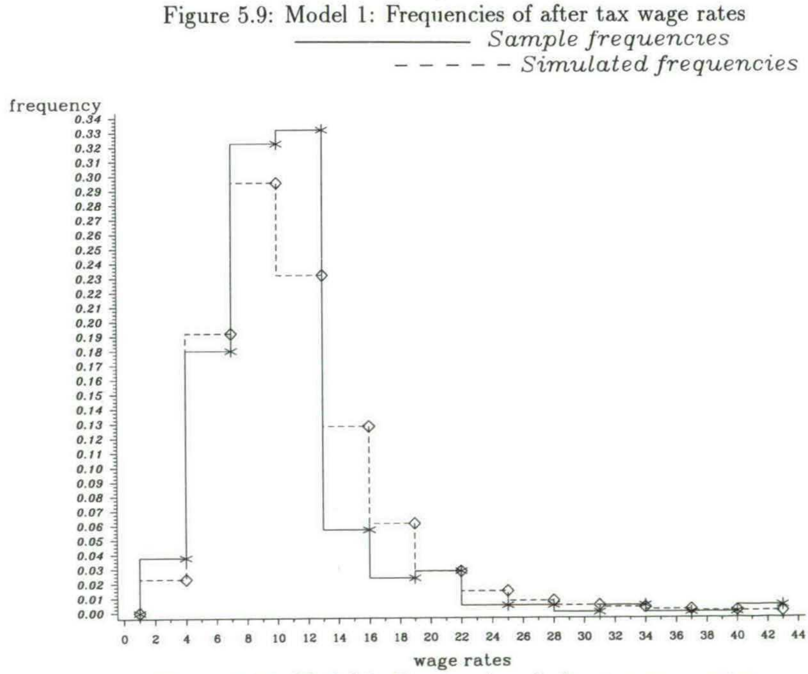

Figure 5.10: Model 2: Frequencies of after tax wage rates

5.5 Conclusions

Two models of job search have been presented. In the first model individuals can determine their labour supply optimally, given the wage rate. The second model assumes that the individual is faced by hours restrictions on the labour market. It has been shown that these two different model assumptions have different implications for the job acceptance decision of individuals. Parameter estimates of the job offer arrival rate show similar effects for both models. Age has a negative effect on the job offer arrival rate. People who live in the industrialized western part of the Netherlands have a larger probability of getting a job offer. Individuals who are not specialized in a certain profession have a lower chance of getting a job offer than other individuals. For individuals who have the Dutch nationality the job offer arrival rate is larger than for those who do not have the Dutch nationality.

For both models the reservation income for every individual has been simulated and compared with the benefit income. For the model without hours restrictions a large percentage of the simulated reservation incomes were below the benefit incomes. This is consistent with results of Van den Berg (1990c) who finds evidence in favour of "disutility of unemployment". For the model with hours restrictions however, most of the reservation incomes are above the benefit incomes. Two explanations for the different results of the two models have been given. The first is that a constrained individual always has a higher (hours dependent) reservation wage than an unconstrained individual, because he has to choose from hours offers that do not coincide with optimal labour supply. The second explanation is that the spike in the sample distribution of weekly working hours at a level of 40 hours a week, together with the assumption of free choice in model 1, lead to low estimates of the virtual wage rate and consequently lowers the lowerbound of the reservation wage rate.

Residual analysis shows that the hazard rate is possibly positive duration dependent, i.e. the longer someone is unemployed, the larger the escape rate is. Positive duration dependence of the hazard rate can be caused by negative duration dependence of the reservation wage rate, possibly due to negative duration dependence of benefit payments or positive business cycle movements. Furthermore, the residual analysis reveals substantial differences between the two model specifications. The model in which the individual can freely choose his working hours seems to be seriously misspecified, whereas the model with hours restrictions is reasonable, in spite of the heavy stationarity assumption.

The simulated frequencies of part time jobs, full time jobs and high level jobs reveal that the poor performance of the model without hours restrictions is mainly caused by the assumption made about labour supply.

5.A Derivation of reservation wage equation 5.7

Let $V(\epsilon)$ denote the value of search of an individual with unobserved taste shifter ϵ. Due to the stationarity assumption (assumption 4 in section 5.2.1), $V(\epsilon)$ is independent of time. At time t, the individual is not working, is looking for a job and receives weekly non-labour income μ and weekly benefits b, which are time independent due to assumption 4 in section 5.2.1. In a short time interval of length Δt the utility flow derived from μ and b equals

$$\int_t^{t+\Delta t} u(b+\mu,0;\epsilon)e^{-\rho(s-t)}ds = \frac{u(b+\mu,0;\epsilon)}{\rho}(1-e^{-\rho\Delta t}) \tag{5.A.1}$$

In the time interval of length Δt there is a probability of $e^{-\lambda\Delta t}\lambda\Delta t + o(\Delta t)$ of receiving a job offer, consisting of a wage rate $\tilde{w}$. The value of the job, denoted by $W(\tilde{w};\epsilon)$ will be compared with the value of continuing searching, which is $V(\epsilon)$. The job offer $\tilde{w}$ will be accepted if $W(\tilde{w};\epsilon)$ exceeds $V(\epsilon)$. Due to the assumption 5 in section 5.2.1, the value will remain $W(\tilde{w};\epsilon)$ once a job $\tilde{w}$ is accepted. With probability $1-e^{-\lambda\Delta t}\lambda\Delta t + o(\Delta t)$ the individual does not get a job offer in the time interval of length Δt, in which case the value remains at $V(\epsilon)$. Summarizing, the value $V(\epsilon)$ is

$$\begin{aligned} V(\epsilon) = u(b+\mu,0;\epsilon)(1-e^{-\rho\Delta t})/\rho + \\ e^{-\rho\Delta t}\{(1-e^{-\lambda\Delta t}\lambda\Delta t)V(\epsilon) + e^{-\lambda\Delta t}\lambda\Delta t E_{\tilde{w}} \max[V(\epsilon), W(\tilde{w};\epsilon)]\} + o(\Delta t) \end{aligned} \tag{5.A.2}$$

Rearranging terms yields

$$\frac{1-e^{-\rho\Delta t}}{\Delta t}V(\epsilon) = \frac{u(b+\mu,0;\epsilon)}{\rho}\frac{(1-e^{-\rho\Delta t})}{\Delta t} + e^{-(\lambda+\rho)\Delta t}\lambda E_{\tilde{w}} \max[0, W(\tilde{w};\epsilon) - V(\epsilon)] + \frac{o(\Delta t)}{\Delta t} \tag{5.A.3}$$

Letting $\Delta t \to 0$ we obtain

$$\rho V(\epsilon) = u(b+\mu,0;\epsilon) + \lambda E_{\tilde{w}} \max[0, W(\tilde{w};\epsilon) - V(\epsilon)] \tag{5.A.4}$$

Using the assumptions 1, 2, 4, 5 and 6 the value of the job with wage rate $\tilde{w}$ can be obtained by solving the maximization problem

$$\max_{c,h} \frac{u(y,h;\epsilon)}{\rho}, \text{ subject to } y = \tilde{w}h + \mu \tag{5.A.5}$$

The solution for h is $h^* = \bar{h}(w,\mu;\epsilon)$ which is the neo-classical labour supply function. Inserting the solution for y and h in the direct utility function yields the indirect utility function which is $\nu(\tilde{w},\mu;\epsilon)$ and therefore

$$W(\tilde{w};\epsilon) = \frac{\nu(\tilde{w},\mu;\epsilon)}{\rho} \tag{5.A.6}$$

Inserting (5.A.6) into (5.A.4) gives

$$\rho V(\epsilon) = u(b+\mu,0;\epsilon) + \frac{\lambda}{\rho}E_{\tilde{w}} \max[0, \nu(\tilde{w},\mu;\epsilon) - \rho V(\epsilon)] \tag{5.A.7}$$

A job offer $\tilde{w}$ will be accepted if $\nu(\tilde{w},\mu;\epsilon) > \rho V(\epsilon)$. If the reverse holds, it will be rejected. As the indirect utility function is increasing in the wage rate, there exists a unique reservation wage rate $\xi(\epsilon)$ such that all wages above it are acceptable and those below will be rejected. The value of search is equal to the value of accepting a job if

$$\rho V(\epsilon) = \nu(\xi(\epsilon),\mu;\epsilon) \tag{5.A.8}$$

Inserting (5.A.8) in (5.A.7) and using the distributional assumptions on $\tilde{w}$ results in the reservation wage equation:

$$\nu(\xi(\epsilon),\mu;\epsilon) = u(b+\mu,0;\epsilon) + \frac{\lambda}{\rho}\int_{\xi(\epsilon)}^{\infty}[\nu(\tilde{w},\mu;\epsilon) - \nu(\xi(\epsilon),\mu;\epsilon)]dF(\tilde{w};\psi,\tau) \tag{5.A.9}$$

5.B Derivation of the reservation utility equation 5.44

Let $V(\epsilon)$ denote the value of search for an individual with unobserved characteristics ϵ. While unemployed the individual receives weekly benefits b. The amount of non-labour income is μ. The flow of utility, derived from benefit level b and non-labour income μ in a short interval of length Δt is, like in (5.A.1), given by:

$$\frac{u(b+\mu,0;\epsilon)}{\rho}(1-e^{-\rho\Delta t}) \tag{5.B.1}$$

The probability of receiving a job offer in a short interval with length Δt is $e^{-\lambda\Delta t}\lambda\Delta t + o(\Delta t)$. A job offer consists of two characteristics i.e. wages and hours which arrive randomly from a joint wage-hours offer distribution. The value of a job with hourly wage rate $\tilde{w}$ and weekly working hours $\tilde{h}$ is denoted by $W(\tilde{w},\tilde{h};\epsilon)$ for an individual with unobserved characteristics ϵ. The value of a job will be compared with the value of search $V(\epsilon)$ in order to make the job acceptance decision. The equivalent of equation (5.A.2) becomes

$$V(\epsilon) = u(b+\mu,0;\epsilon)(1-e^{-\rho\Delta t})/\rho + \tag{5.B.2}$$
$$e^{-\rho\Delta t}\{(1-e^{-\lambda\Delta t}\lambda\Delta t)V(\epsilon) + e^{-\lambda\Delta t}\lambda\Delta t E_{(\tilde{w},\tilde{h})}\max[V(\epsilon), W(\tilde{w},\tilde{h};\epsilon)]\} + o(\Delta t)$$

After rearranging terms and taking the limit $\Delta t \to 0$ we get

$$\rho V(\epsilon) = u(b+\mu,0;\epsilon) + \lambda E_{(\tilde{w},\tilde{h})}\max[0, W(\tilde{w},\tilde{h};\epsilon) - V(\epsilon)] \tag{5.B.3}$$

The assumptions 2 and 5 are used to determine the value of the job with wage rate $\tilde{w}$ and hours $\tilde{h}$:

$$W(\tilde{w},\tilde{h};\epsilon) = \frac{u(\tilde{w}\tilde{h}+\mu,\tilde{h};\epsilon)}{\rho} \tag{5.B.4}$$

Inserting this expression in (5.B.3) yields

$$\rho V(\epsilon) = u(b+\mu,0;\epsilon) + \frac{\lambda}{\rho}E_{(\tilde{w},\tilde{h})}\max[0, u(\tilde{w}\tilde{h}+\mu,\tilde{h};\epsilon) - \rho V(\epsilon)] \tag{5.B.5}$$

A job offer $(\tilde{w},\tilde{h})$ is accepted if the utility value $u(\tilde{w}\tilde{h}+\mu,\tilde{h};\epsilon)$ exceeds $\rho V(\epsilon)$. In other words, there exists an unique reservation utility level $\bar{u}(\epsilon)$ with

$$\bar{u}(\epsilon) = \rho V(\epsilon) \tag{5.B.6}$$

All job offers which generate a utility level above $\bar{u}(\epsilon)$ are acceptable, those below will be rejected. According to assumption 6, $u(\tilde{w}\tilde{h}+\mu,\tilde{h};\epsilon)$ is increasing in $\tilde{w}$, and therefore, for a given level of $\tilde{h}$, a wage rate $\xi(\tilde{h},\bar{u}(\epsilon);\epsilon)$ can be found such that

$$u(\xi(\tilde{h},\bar{u}(\epsilon);\epsilon)\tilde{h}+\mu,\tilde{h};\epsilon) = \bar{u}(\epsilon) \tag{5.B.7}$$

This $\xi(\tilde{h},\bar{u}(\epsilon);\epsilon)$ can be interpreted as an hours dependent reservation wage rate. Using the assumption that wage and hours arrive simultaneously from a joint wage-hours offer distribution $f(\tilde{w},\tilde{h})$, equation (5.B.5) can be rewritten such that the reservation utility equation (5.44) is obtained:

$$\bar{u}(\epsilon) = u(b+\mu,0;\epsilon) + \frac{\lambda}{\rho}\int_0^\infty\int_{\xi(\tilde{h},\bar{u}(\epsilon);\epsilon)}^\infty [u(\tilde{w}\tilde{h}+\mu,\tilde{h};\epsilon) - \bar{u}(\epsilon)]f(\tilde{w},\tilde{h})d\tilde{w}d\tilde{h} \tag{5.B.8}$$

5.C Derivation of (5.35)

In this appendix we show that if the derivatives of the vector of scores with respect to the implicitly defined integration bounds are replaced by (5.35) the expectation of the resulting simulator still equals zero in the true parameter vector. In order to write down the complete density function some dummy variables are introduced.

$$\begin{aligned} d &= 1 \quad \text{for incomplete spells } (t > M) && (5.C.1)\\ &= 0 \quad \text{for completed spells } (t \leq M) && (5.C.2)\\ e &= 1 \quad \text{if } d = 0 \text{ and } (w,h) \text{ is observed} && (5.C.3)\\ &= 0 \quad \text{if } d = 0 \text{ and } (w,h) \text{ is unobserved} && (5.C.4)\end{aligned}$$

The density function consists of the following parts:

$$\begin{aligned} l_u(\eta) &\quad \text{if} \quad d = 1 && (5.C.5)\\ (1-\pi)l_{cu}(\eta) &\quad \text{if} \quad d = 0, e = 0 && (5.C.6)\\ \pi l_{co}(\eta) &\quad \text{if} \quad d = 0, e = 1 && (5.C.7)\end{aligned}$$

where $\pi = E(e)$, the probability that the post unemployment job characteristics are observed which has been assumed to be independent of the parameters of interest and therefore has been neglected before. $l_u(\eta)$, $l_{cu}(\eta)$ and $l_{co}(\eta)$ have been defined in (5.12), (5.21) and (5.19) respectively. To show that the density integrates to 1 we look at each contribution separately.

$$\int_0^M l_{cu}(\eta)dt = 1 - \int_{-\infty}^{\infty} \exp\{-\theta(\epsilon)M\}g(\epsilon,\sigma_\epsilon^2)d\epsilon = 1 - l_u(\eta) \tag{5.C.8}$$

$$\int_0^M \int_0^\infty \int_0^\infty l_{co}(\eta)dwdhdt = \tag{5.C.9}$$

$$\int_0^M \int_0^\infty \int_0^\infty \int_{I_w} \theta(\epsilon)\exp\{-\theta(\epsilon)t\}r(h|w,\epsilon)\frac{f(w)}{T(\epsilon)}g(\epsilon,\sigma_\epsilon^2)d\epsilon dwdhdt = \tag{5.C.10}$$

$$\int_0^M \int_0^\infty \int_{-\infty}^\infty \int_{\xi(\epsilon)}^\infty \theta(\epsilon)\exp\{-\theta(\epsilon)t\}r(h|w,\epsilon)\frac{f(w)}{T(\epsilon)}g(\epsilon,\sigma_\epsilon^2)dwd\epsilon dhdt = \tag{5.C.11}$$

$$1 - l_u(\eta) \tag{5.C.12}$$

From these results it can be derived that

$$l_u(\eta) + (1-\pi)\int_0^M l_{cu}(\eta)dt + \pi\int_0^M \int_0^\infty \int_0^\infty l_{co}(\eta)dwdhdt = 1 \tag{5.C.13}$$

As a consequence

$$\frac{\partial l_u(\eta)}{\partial\eta} + (1-\pi)\int_0^M \frac{\partial l_{cu}(\eta)}{\partial\eta}dt + \pi\int_0^M \int_0^\infty \int_0^\infty \frac{\partial l_{co}(\eta)}{\partial\eta}dwdhdt = 0 \tag{5.C.14}$$

or, equivalently,

$$\frac{\partial l_u(\eta)}{\partial\eta} + (1-\pi)\int_0^M \frac{\partial l_{cu}(\eta)}{\partial\eta}dt + \pi\frac{\partial\left[\int_0^M \int_0^\infty \int_0^\infty l_{co}(\eta)dwdhdt\right]}{\partial\eta} = 0 \tag{5.C.15}$$

as

$$\int_0^M \int_0^\infty \int_0^\infty \frac{\partial l_{co}(\eta)}{\partial \eta} dwdhdt = \frac{\partial \left[\int_0^M \int_0^\infty \int_0^\infty l_{co}(\eta) dwdhdt \right]}{\partial \eta} \tag{5.C.16}$$

Now, note that the left hand side of (5.C.16) is equal to the expectation of $e\partial \ln l_{co}(\eta)/\partial \eta$, the score contribution which we want to simulate. The derivatives of (5.C.16) consist of two major terms, i.e. the derivatives with respect to the bounds and the derivatives with respect to the integrand. As the derivatives with respect to the integrand of the right hand side are equal to the derivatives with respect to the bounds on the left hand side, it follows that the derivatives with respect to the bounds are equal on both sides as well. Using the right hand side of (5.C.16) and the fact that (5.C.9) equals (5.C.11) the derivatives with respect to the bounds can be written as

$$-\int_{-\infty}^{\infty} \frac{\partial \xi(\epsilon)}{\partial \eta} [1 - \exp\{-\theta(\epsilon) M\}] \frac{f(\xi(\epsilon))}{T(\xi(\epsilon))} g(\epsilon, \sigma_\epsilon^2) d\epsilon \tag{5.C.17}$$

which is expression (5.35). To handle the problem with the implicitly defined bounds, the original derivatives are replaced by their expectation.

Chapter 6

Job search, search intensity and labour market transitions: an empirical exercise

6.1 Introduction

In this chapter an empirical model of job search is presented in which the intensity of search is determined endogenously by the individual decision maker. Individuals decide on both the amount of effort that they spend on search activities and on whether or not to accept a job offer, as opposed to the standard job search model, see e.g. McKenna (1985), in which the emphasis is on the job acceptance decision only.

In the literature, various models of job search with endogenously determined intensity of search have appeared. Burdett and Mortensen (1978) present a model in which the total time endowment is divided in time spent on labour, leisure time and time spent on search. By increasing the time spent on search individuals can increase the average number of job offers arriving in a given time interval. At the same time, however, they have to incur a utility loss because an increase in time spent on search implies a decrease in leisure time. Individuals determine the optimal amount of time spent on search by determining the optimal trade-off between the returns in the form of expected job offers and the cost due to the loss of leisure time. This trade-off between cost of search and returns of search is common to all models on search intensity that have appeared in literature. In Mortsenen (1986) a simpler version of the same model is presented. Instead of expressing search intensity in terms of time spent on search, he defines search intensity much more loosely in terms of "search effort". An explicit cost of search function is formulated and again an increase in search effort raises the job offer arrival rate. Yoon (1981) also specifies a cost of search function. He estimates a model of unemployment duration in which a measure of search intensity ("The number of places looked for a job in the past few weeks") appears as an explanatory variable in the hazard rate and he finds a significant negative effect, i.e. an increase in the intensity of search tends to reduce the duration of unemployment. Lindeboom and Theeuwes (1992) specify a reduced form model in which data on the number of search contacts appears in the hazard rate. The emphasis in their study is on detecting the relation between the intensity of search and elapsed duration.

Point of departure of the model of this chapter is the search model by Mortensen (1986). In this model, both on the job searchers and searchers without a job are considered. To make the model suitable for empirical application some of Mortensen's simplifying assumptions have to be relaxed. We allow for differences in cost of search for workers and non-workers and for non-zero cost of turnover. Moreover, differences in the job offer arrival rate between different labour force states may occur and individuals may prefer a certain labour force state, other things being equal. Optimal search intensity is chosen such that the marginal cost of search is equal to the marginal returns, conditional on search intensity being non-negative. An optimal search intensity of zero may arise as a corner solution. The expected marginal returns of search depend on the labour market conditions, as well as on the present earnings and expected future earnings. It can be shown that there is a negative relationship between present earnings and the optimal intensity of search. For an employed individual this means that if the present wage income is above a certain threshold value, he will decide not to search. This implies that there is jointness in the decision whether or not to search and the distribution of observed wage income. This relation can be examined using cross section data in a limited dependent variables model. Data on search duration can be used to estimate the relation between search intensity and duration. The data we use are from the Dutch Socio Economic Panel. From October 1987 on questions concerning the search behaviour of individuals appeared in the survey. Instead of having one measure for search intensity, the data set contains several indicators for search intensity that are related to different search instruments. Therefore, "search intensity" in the structural model is allowed to be a vector of search instruments on which the individual decides.

In section 6.2, the economic model is presented. In section 6.3, various empirical specifications are presented. First of all, a reduced form model is specified in which the employment decision, the search decision and the wage distribution are modelled jointly. After that, a structural duration model is specified which allows for transitions from unemployment into employment, turnover transitions and transitions from employment into unemployment. Cost of search functions are specified and differences in cost of search for different labour market states are examined. The duration model is estimated jointly with the structural equation for search intensity and parameter estimates for the job offer arrival rates and the cost of search functions are obtained. The estimation results are presented. The final section concludes.

6.2 The economic model

The economic model is based on the model of Mortensen (1986). We extend the model in order to make it more realistic. For example, we allow the cost of search functions and the job offer arrival rates to be different for employed and unemployed individuals. The within period utility level may be different for different labour force states and we allow for the presence of non-zero cost of turnover.

It is assumed that individuals maximize their lifetime utility, taking into account uncertainty about their future labour force state and taking into account that they can influence the expected number of job offers by searching more or less intensely. The within period utility function depends on the level of income in the given period. Given the level of income, there may be differences in the utility level depending on the labour force state of the individual. The same specification for within period utility was used by Narendranathan and Nickell (1985) and Van den Berg (1990). In their application they found evidence in favour of "disutility of unemployment", i.e. $\omega < 1$.

$$\begin{aligned} \text{utility(income=x,state=employed)} &= u(x) \\ \text{utility(income=x,state=unemployed)} &= \omega u(x) \end{aligned} \tag{6.1}$$

with $u'(x) > 0$.

For the unemployed, job offers arrive according to a Poisson process with parameter $(1+\alpha_u s)\lambda_u$, which is a function of search intensity s, $s \geq 0$, that can be determined by the individual himself, λ_u is determined by the demand side of the labour market, and $\alpha_u \geq 0$ is a parameter that determines the effectiveness of search. Note, that at an intensity of search of zero, the job offer arrival rate reduces to λ_u, leaving open the possibility of getting a job offered without search. As our dataset contains various indicators for the intensity of search which provide information on the use of different search instruments, search intensity s, and consequently α_u, is allowed to be a vector of different search instruments. For ease of exposition, search intensity is treated as a one dimensional variable in this section, but the extension to more dimensions is straightforward.

For employed individuals the job offer arrival rate is indicated by $(1 + \alpha_e s)\lambda_e$.

The cost of search is a function of search intensity, indicated by $c_u(s)$ for unemployed individuals and $c_e(s)$ for employed individuals. The cost of search function is assumed to have the following properties:

$$\begin{aligned} c_x(0) &= 0 \\ c_x'(s) &> 0 \\ c_x''(s) &> 0 \\ x &= e, u \end{aligned} \tag{6.2}$$

The first condition simply states, no search, no cost, the second condition implies more search, higher cost and the third is a regularity condition on the cost function in order to ensure that an optimal value of search intensity exists.

An employed individual who changes his job is faced by a cost of turnover of k.[1] A job offer is modelled as a random draw from a wage distribution $F(.)$, which is assumed to be known to the individual. For employed individuals there is an exogenous layoff rate σ. The subjective rate of discount is denoted by ρ.

The value function for an unemployed individual, which is the maximum of the expected lifetime utility, is indicated by V. For an employed individual who earns a wage income of w, the value function is indicated by $W(w)$. Below the Bellman equations are

[1] Van den Berg (1992) formulated a model of job to job transitions in which he made the cost of turnover a function of the present wage.

presented. Their derivation is given in the appendix.

$$\begin{array}{l}\rho V = \max_{s\geq 0}\{\omega u(b+\mu) - c_u(s) + (1+\alpha_u s)\lambda_u \int_0^\infty \{\max[V, W(x)] - V\} dF(x)\} \\ (\rho+\sigma)W(w) = \\ \max_{s\geq 0}\{u(w+\mu) - c_e(s) \\ +(1+\alpha_e s)\lambda_e \int_0^\infty \{\max[V, W(x)-k, W(w)] - W(w)\} dF(x) + \sigma V\}\end{array} \tag{6.3}$$

in which b is benefit income and μ is non-labour income, both assumed to be fixed over time (stationarity assumption). The first equation is for unemployed individuals. It is equal to the within period utility evaluated at the present income $b+\mu$, minus the cost of search, evaluated at the optimal intensity of search, plus the expected returns of search. If an unemployed individual gets a job offer with wage x, he compares the value of being employed at wage x, $W(x)$, with the value of being unemployed V and he chooses the alternative for which the value function has the highest level. This implies that the reservation wage ξ for an unemployed individual is implicitly defined by $V = W(\xi)$[2]. The reservation wage income is the wage income for which the individual is indifferent between working and not working. For an employed individual the same type of Bellman equation holds.

The value function for an employed individual, who is currently employed at wage w, is equal to the discounted value of the within period utility, evaluated at the present income, minus cost of search plus expected returns of search. Note that the consequence of a nonzero layoff rate σ is that there are different discount rates for employed and unemployed individuals, i.e. for the employed the discount rate is increased by σ. An employed individual who gets a job offer with wage x has three choices: he can stop working, he can accept the job, in which case he is faced by a cost of turnover of k, or he can keep his present job. Actually, the first alternative is irrelevant, for if V exceeded $W(w)$, he would not be observed to be working. For an employed individual we can define a reservation wage $\alpha(w)$, which is the wage at which he is indifferent between his present job and the job offer, i.e. $W(\alpha(w)) = W(w)+k$. Note that if the cost of turnover is zero, the individual will accept any job that has a higher wage than the present job. As $W'(.) > 0$, $\alpha(w)$ will exceed w as long as the cost of turnover k is positive.

From the Bellman equations the first order conditions for optimal search intensity can be derived:

$$\begin{array}{rcl} c_u'(\bar{s}_u) & = & \alpha_u\lambda_u \int_\xi^\infty [W(x) - V] dF(x) \\ s_u^* & = & \max\{0, \bar{s}_u\} \\ c_e'(\bar{s}_e(w)) & = & \alpha_e\lambda_e \int_{\alpha(w)}^\infty [W(x) - W(\alpha(w))] dF(x) \\ s_e^*(w) & = & \max\{0, \bar{s}_e(w)\} \end{array} \tag{6.4}$$

where use has been made of $W(\alpha(w)) = W(w) + k$ and $V = W(\xi)$. s_u^* and $s_e^*(w)$ are the optimal intensities of search for unemployed individuals and employed individuals at wage w respectively. For positive values of search intensity the marginal cost of search are equal to the marginal returns of search. If the solution of the marginal cost = marginal returns equation is nonpositive, optimal search intensity will be zero.

[2] $W(.)$ is an increasing function of the wage. This can be proved by contradiction: Suppose that $W(.)$ is non-increasing in the wage: then the right hand side of (6.3) is increasing in the wage. The property that $W(w)$ is increasing in w is sufficient to guarantee the existence of a unique reservation wage, both for unemployed and employed individuals

Figure 6.1 depicts the cost of search and the returns of search as a function of search intensity for an unemployed individual. The returns of search are a straight line. The gains of search, i.e. the difference between returns and cost are maximal if the slopes are equal, which is at $\bar{s}_u$ in the picture. In this particular case $\bar{s}_u$ is positive and consequently the optimal search intensity s_u^* is positive. If the marginal returns of search rise due to an increase in the effectiveness of search parameter α_u, the intensity of search at which marginal cost of search and marginal returns of search are equal will rise as well. This is because, due to conditions (6.2), the marginal cost of search rises with search intensity. The higher marginal returns of search are indicated by the dotted line, and the point at which slopes are equal is $\bar{\bar{s}}_u$. In figure 6.2 the situation is depicted in which it is optimal not to search. Here the gains of search at any positive intensity of search are lower than the gains of search at a search intensity of zero: The marginal cost of search are higher than the marginal returns of search at a zero search intensity.

According to equations (6.4) and figures 6.1 and 6.2, there is a positive relationship between search intensity and marginal returns of search. Given that marginal cost of search are positive, this relation is implied by the condition in (6.2) that states that $c_x"(s) > 0$. In the empirical implementation, this is a testable implication. The linearity of the returns of search, together with the curved cost of search function, guarantee that the marginal cost = marginal returns equation has a unique solution, and uniqueness is what we need to arrive at an estimable relationship. However, linearity of the returns of search is by no means necessary for a unique solution, and the implied positive relationship between optimal search intensity and marginal returns of search does not depend on these shapes of the cost and returns of search functions either. If we choose a returns of search specification with decreasing marginal returns, together with a cost of search specification that satisfies (6.2), uniqueness and the positive relationship between search intensity and marginal returns of search are maintained. The same holds for decreasing marginal returns and a linear cost function. To test for the positive relationship between marginal returns and search intensity, we need either a linear returns of search function and a curved cost of search function, or a linear cost of search function and a curved returns of search function, otherwise there is no unique solution (i.e. no solution or multiple solutions) for the marginal cost = marginal returns equation if the positive relationship is <u>not</u> satisfied. Given the fact that our information on search intensity, that is contained in the dataset, partly consists of qualitative variables (binary indicators) of which scale is not identified, it makes little sense to test for the presence of decreasing marginal returns of search. Therefore, we maintain the linear returns specification and the curved cost of search function. The expression for the marginal returns of search that we find at the right hand side of (6.4) will depend on the specific functional forms that are chosen. However, irrespective of the chosen functional forms, marginal returns will be decreasing in the reservation wage. As the reservation wage depends, among other things, on current income (benefit income for unemployed individuals and wage income for employed individuals) this implies that there is always a negative relation between current income and marginal returns, irrespective of the functional forms chosen.

As the reservation wage is an important determinant of the marginal returns, we now will consider the reservation wages and their relationship with search intensity in more detail.

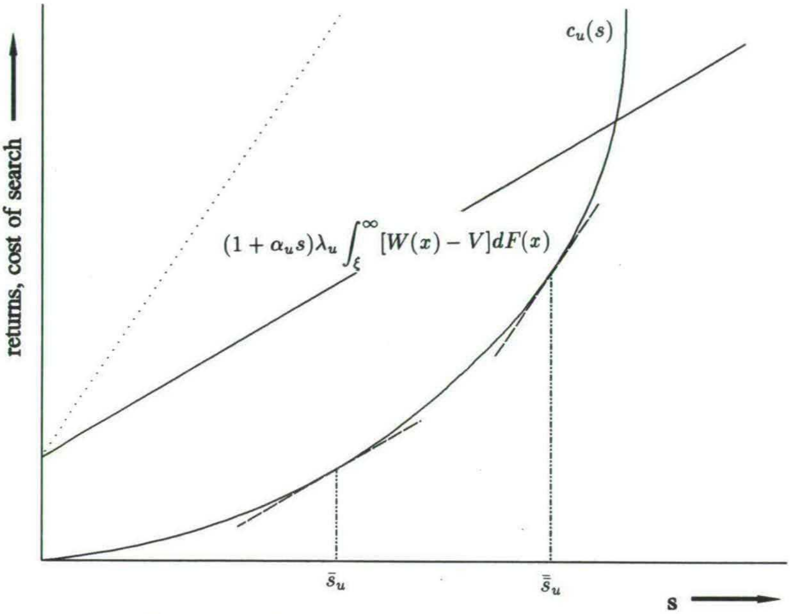

Figure 6.1: Determination of optimal search intensity

Above we have introduced the reservation wages, ξ and $\alpha(w)$ for unemployed and employed individuals respectively, which characterize the job acceptance decision, i.e. a job offer with a wage above the reservation wage rate will be accepted. In a similar way, we can define a so called search reservation wage y_e for employed individuals and y_u for unemployed individuals. The search reservation wage is the income value at which the individual is indifferent between searching or not, i.e. it is the income value for which marginal cost of search and marginal returns of search at a search intensity of zero are equal:

$$\alpha_z\lambda_z \int_{y_z}^{\infty}[W(x)-W(y_z)]dF(x) = c_z'(0), z=e,u \tag{6.5}$$

Note from (6.4) that the marginal returns of search are decreasing in ξ and $\alpha(w)$ for unemployed and employed individuals respectively. Consequently, if $\xi > y_u$ the marginal returns of search will be smaller than the marginal cost of search at a zero search intensity and it will be optimal not to search. If, in contrast, $\xi < y_u$, the individual can increase expected gains by choosing a positive intensity of search. The same holds for employed individuals for whom we have to compare $\alpha(w)$ and y_e.

The following relations can be defined:

$$\begin{aligned} s_u^* > 0 &\Leftrightarrow y_u > \xi \\ s_e^*(w) > 0 &\Leftrightarrow y_e > \alpha(w) \\ s_u^* = 0 &\Leftrightarrow y_u \leq \xi \\ s_e(w) = 0 &\Leftrightarrow y_e \leq \alpha(w) \end{aligned} \tag{6.6}$$

As $\alpha'(.)$ is a positive function of wage income, there is a direct relation between the

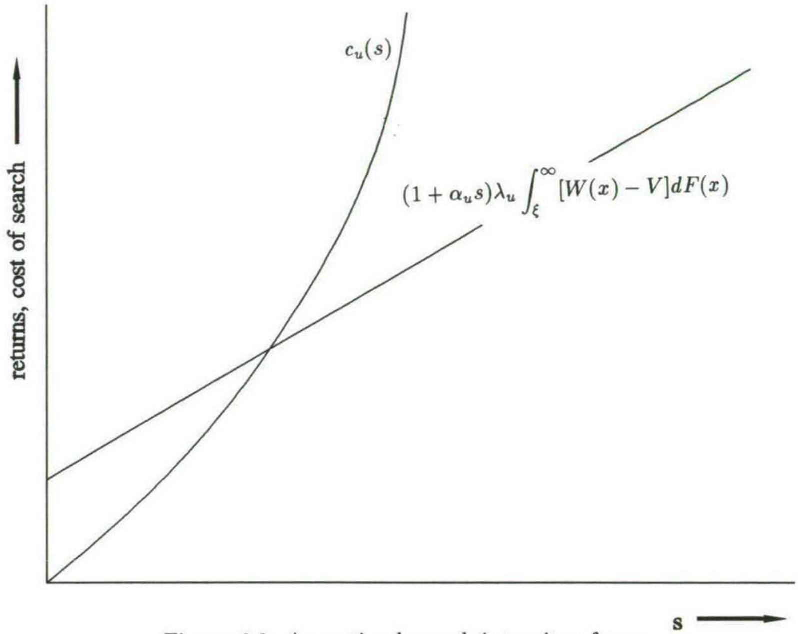

Figure 6.2: An optimal search intensity of zero

current wage income and the decision of whether or not to search. As the reservation wage $\alpha(w)$ for employed rises with turnover cost k, higher cost of turnover implies lower marginal returns to search and therefore a decreased incentive to search.

For the employed individual we have seen that there is a negative relation between the present wage and the intensity of search. To say something about the determinants of the reservation wage ξ of unemployed individuals we write the Bellman equations in a simpler form:

$$\begin{array}{lcl} \rho V & = & \omega u(b+\mu) - c_u(s_u^*) + (1+\alpha_u s_u^*)\lambda_u \int_{\xi}^{\infty}[W(x) - W(\xi)]dF(x) \\ (\rho+\sigma)W(w) & = & u(w+\mu) - c_e(s_e^*(w)) \\ & & +(1+\alpha_e s_e^*(w))\lambda_e \int_{\alpha(w)}^{\infty}[(W(x) - W(\alpha(w))]dF(x) + \sigma V \end{array} \tag{6.7}$$

The reservation wage ξ is implicitly defined by $V = W(\xi)$ or

$$\begin{array}{c} \omega u(b+\mu) - c_u(s_u^*) + (1+\alpha_u s_u^*)\lambda_u \int_{\xi}^{\infty}[W(x) - W(\xi)]dF(x) = \\ u(\xi+\mu) - c_e(s_e^*(\xi)) + (1+\alpha_e s_e^*(\xi))\lambda_e \int_{\alpha(\xi)}^{\infty}[W(x) - W(\alpha(\xi))]dF(x) \end{array} \tag{6.8}$$

From this equation, little insight is gained at first sight and therefore we start by looking at some special cases. Suppose that $\sigma = 0, \omega = 1, \lambda_e = \lambda_u, k = 0, \alpha_e = \alpha_u$ and $c_u(.) \equiv c_e(.)$. This is the standard case treated by Mortsensen (1986). It can easily be seen that under these conditions $s_u^* = s_e^*(\xi)$ and $\xi = b$. That is, in the absence of any differences in the cost functions and job offer arrival rates, if cost of turnover are zero and if individuals have no preference for a certain labour force state, the reservation income ξ is equal to the benefit income. Any job offer with a wage income that is only one guilder higher than the benefit income will be accepted by the individual, knowing that once he has got a job, he can continue searching under the same conditions as when he was unemployed. Therefore, cost of search and expected future returns have no influence on the reservation income, which is fully determined by current income. For the decision whether to search or not this means that if the benefit income is above the search reservation income y_u the unemployed individual will not search at all.

Now we relax the restriction $\omega = 1$, but the other restrictions remain active. Then the reservation income is defined by $u(\xi + \mu) = \omega u(b + \mu)$. If the individual prefers being unemployed over being employed at the same income, ω exceeds one and as a consequence $\xi > b$. Now there is a certain threshold. The individual will not simply accept any job with a wage that is only one cent higher than the benefit income. He needs to be compensated for the loss in utility he incurs by changing to the less preferable state of employment. The reverse holds for the case of disutility of unemployment, $\omega < 1$, in which $\xi < b$. Note, that the reservation wage is still determined by current period characteristics as in both states still the same search conditions hold.

Now we relax the assumptions $\lambda_e = \lambda_u, c_u(.) \equiv c_e(.)$ and $k = 0$, allowing for differences in the arrival rates and the cost functions. Then the reservation wage is defined by:

$$\begin{aligned} u(\xi+\mu) - \omega u(b+\mu) = & \left[(1+\alpha_u s_u^*)\lambda_u \textstyle\int_\xi^\infty [W(x) - W(\xi)]dF(x) - c_u(s_u^*)\right] \\ & - \left[(1+\alpha_e s_e^*(\xi))\lambda_e \textstyle\int_{\alpha(\xi)}^\infty [W(x) - W(\alpha(\xi))]dF(x) - c_e(s_e^*(\xi))\right] \end{aligned} \tag{6.9}$$

Note, that both sides are zero if there are no differences in the arrival rates and cost functions and if turnover costs are zero. On the right hand side we see the difference of the expected gains of search (returns minus costs) while unemployed and the expected gains of search while employed at the reservation wage. If the expected gains of search while unemployed are higher than the expected gains of search while employed, the reservation wage rate ξ will be higher than in the restricted case that we described just before. The individual is less eager to accept a job, knowing that once he has accepted a job he enters a state in which the conditions of search are less favourable. The reverse holds for the case in which the expected gains of search while employed are higher than the expected gains of search while unemployed. Then we find the reservation wage to be lower than in the restricted case. Note that as higher cost of turnover tends to reduce the gains of search while employed, the reservation wage ξ for unemployed will be higher as well. Finally, note that a positive layoff rate σ has no consequences for the reservation wage ξ. A positive layoff rate decreases the value of employment, but the value of unemployment is decreased as well as the returns of search while unemployed depend on the expected value of employment.

Until now the reservation wage rate $\alpha(w)$ for employed individuals has only been defined implicitly by the relation $W(\alpha(w)) = W(w) + k$. From the defining relation it can be derived that

$$\alpha'(w) = \frac{u'(w+\mu)}{u'(\alpha(w)+\mu)} \frac{\rho + \sigma + (1+\alpha_e s_e^*(\alpha(w)))\lambda_e \bar{F}(\alpha(\alpha(w)))}{\rho + \sigma + (1+\alpha_e s_e^*(w))\lambda_e \bar{F}(\alpha(w))} \tag{6.10}$$

It is not straightforward to solve this differential equation. Van den Berg (1992) proposes to use a Taylor expansion around zero costs of turnover, making use of the relation

$\alpha(w) = w \Leftrightarrow k = 0$. In the present context this Taylor expansion is

$$\alpha(w) = w + \frac{\rho + \sigma + (1 + \alpha_e s_e^{**}(w))\lambda_e \bar{F}(w)}{u'(w + \mu)} k + o(k) \tag{6.11}$$

with $s_e^{**} = \max\{0, \bar{s}_e^*(w)\}$,

$$c_e'(\bar{s}_e^*(w)) = \alpha_e \lambda_e \int_w^\infty [W(x) - W(w)] dF(x) \tag{6.12}$$

6.3 Empirical specification

6.3.1 A reduced form specification for jointly modelling the search decision and the wage equation

From the analysis in the preceding section it has become clear that there is a one to one relation between the current wage and the decision whether or not to search. The higher the current wage income is, the less likely it will be that an individual will search for a job. Apart from the relation between the search decision and the distribution of observed wages, there is a relation between the employment decision and the distribution of observed wages. Only wage offers that are higher than the reservation wage rate are observed. To make inference with respect to the parameters of the wage offer distribution, we have to take into account that there could be correlation between the observed wages, the employment decision and the search decision. In this section, a reduced form wage-search-employment model is estimated. The estimates of the wage equation, obtained in this subsection, can be used in the estimation of the structural model of unemployment duration and job tenure that is described in section 6.3.2.

Before the model is specified, some sample statistics are presented. The data are from the Dutch Socio Economic Panel. The first wave in the panel in which information on individual search behaviour was gathered is that of October 1987, and it is this particular wave that we use to look at the relation between wages, employment and search.

First of all the sample is restricted to individuals who are either employed or unemployed. This selection has been made on the basis of occupational variables in the sample: every individual is asked to report his or her occupation in each month of the past six months, i.e. they are asked whether they were employed, or in full time education, in the forces, disabled, unemployed etc. After having selected the employed and unemployed individuals, we looked at the questions about search. The first question is *"Are you searching for a paid job at the moment, or if you already have a paid job, are you searching for a different one?"* The answers that the respondent can give are *"Yes, I am searching seriously"*, *"Yes, I am thinking about it"*, and *"No"*. The two different possibilities to answer *"yes"* provide two different positive levels of search. If the respondent has answered positively to this first question, some additional questions have to be answered. The first one is *"Have you been looking for work in the past two months?"*, to which the respondent can answer yes or no. By "looking for work" in this context is meant responding to an advertisement, placing an advertisement, gaining information from employers, relatives or the employment office, screening the advertisements etc.

The second question is *"How many times have you applied for a job in the past two months?"* By "applying for a job" is meant writing a letter of application, making a phone call, etc. The final question which gains information about the intensity of search is *"Are you registered at the employment office?"* For an individual with a positive intensity of search four indicator variables $\hat{s}_1, \hat{s}_2, \hat{s}_3$ and $\hat{s}_4$ can be constructed:

$$
\begin{aligned}
\hat{s}_1 &= 1 && \text{if searching seriously} \\
&= 0 && \text{if not} \\
\hat{s}_2 &= 1 && \text{if looking for work in the past two months} \\
&= 0 && \text{if not} \\
\hat{s}_3 &= 1 && \text{if registered at the employment office} \\
&= 0 && \text{if not} \\
\hat{s}_4 &= && \text{number of applications in the past two months}
\end{aligned}
$$

These four variables are indicators for the unobserved intensity of search. In the next subsection, the unobserved intensity of search is modelled by a latent variable, after which the relation between the latent variable and the indicators defined here is specified. For the purpose of the present section, it is sufficient to know whether one is searching or not. Respondents who answered the first question negatively are treated as non-searchers. Moreover, individuals who answered "Yes, I'm thinking about it" to this first question, but who neither looked for a job, nor applied for a job in the past two months, nor registered themselves at the employment office, are treated as non-searchers as well. All individuals for whom at least one indicator $\hat{s}_j$ is positive are classified as searchers. The sample consists of 3016 male individuals. Table 6.1 shows the percentages of searchers, nonsearchers and employed, unemployed in the sample. We see that 12.4% of the employed individuals is searching, whereas for the unemployed individuals the percentage is much higher, namely 79.7. In table 6.2 we report the means of various weekly after tax income variables. Positive non-labour income is the mean of non-labour income restricted to those individuals who have non-zero non-labour income. The same holds for positive benefit income. In the context of the theoretical model, the differences in income between employed and unemployed individuals can be an explanation for the fact that many more unemployed individuals than employed individuals are searchers. There is a considerable difference between the mean wage of employed searchers and the mean wage of employed non-searchers. The mean wage of searchers is lower than the mean wage of non-searchers, which is in accordance with the theoretical model in the previous section, which predicts a negative relation between the current wage and the decision to search. Moreover, the non-labour income of searchers seems to be lower than that of non-searchers, which may be another incentive for them to search. For the unemployed individuals there is not much difference between the benefit income of searchers and non-searchers. If we restrict ourselves to individuals with a positive benefit income, the mean benefit income for non-searchers is even lower than for searchers. Moreover, non-labour income for non-searchers is lower than for searchers. At first sight this seems to be in contrast with the theory in the preceding section in which benefit income was shown to be an important determinant of the reservation wage rate, and consequently an important determinant of the search decision in (6.6), i.e. given everything else, the

Table 6.1 Sample Statistics, October 1987

% employed	:	92.6
% unemployed	:	7.4
% searchers	:	17.4
% non-searchers	:	82.6
searchers		
%unemployed	:	33.8
% employed	:	66.2
non-searchers		
%unemployed	:	1.8
% employed	:	98.2
employed		
% searchers	:	12.4
% non-searchers	:	87.6
unemployed		
% searchers	:	79.7
% non-searchers	:	20.3

higher the benefit income, the less likely it is that one searches. There is, however, an alternative explanation for the fact that non-searching individuals tend to have lower income levels than searching individuals. The fact that these unemployed non-searchers have a low mean non-labour income may be due to the fact that these individuals have been unemployed for quite a long time and therefore they have been using up their assets. Having been unemployed for a long time implies that their expected returns of search are probably low, i.e. their λ_u is low. From (6.5) it can be derived that a low value of λ_u, given the marginal cost, leads to a low value of y_u, which is the threshold value for the reservation wage, above which no search will take place. So even if the individuals have low income variables, they can be observed to be non-searching because low marginal returns of search lead to a low threshold y_u. Now the wage-search-employment model will be specified. It is assumed that wages are lognormally distributed. The model is:

$$\ln w = \eta' x + v \tag{6.13}$$
$$y_1^* = \alpha' z_1 + u_1 \tag{6.14}$$
$$y_2^* = \beta' z_2 + u_2 \tag{6.15}$$

in which the error terms are jointly normally distributed with mean zero and covariance matrix Σ:

$$\Sigma = \begin{pmatrix} \sigma_v^2 & \sigma_{vu_1} & \sigma_{vu_2} \\ \sigma_{vu_1} & 1 & \sigma_{u_1 u_2} \\ \sigma_{vu_2} & \sigma_{u_1 u_2} & 1 \end{pmatrix} \tag{6.16}$$

Table 6.2 Means of the weekly income variables

employed		
wage	:	620.81
wage of searchers	:	533.45
wage of non-searchers	:	633.19
non-labour income	:	18.49
non-labour income searchers	:	14.85
non-labour income non-searchers	:	19.01
positive non-labour income	:	54.58
positive non-labour income searchers	:	44.43
positive non-labour income non-searchers	:	56.00
unemployed		
benefit income	:	203.26
benefit income searchers	:	202.87
benefit income non-searchers	:	204.76
non-labour income	:	41.07
non-labour income searchers	:	45.69
non-labour income non-searchers	:	22.91
positive benefit income	:	286.42
positive benefit income searchers	:	288.38
positive benefit income non-searchers	:	279.22
positive non-labour income	:	168.85
positive non-labour income searchers	:	179.71
positive non-labour income non-searchers	:	114.57

Table 6.3 The wage equation

	estimate	standard error
η_1 const	-10.372**	0.889
η_2 log(age)	9.018**	0.500
η_3 square of log(age)	-1.188**	0.0701
η_4 educ1	-0.476**	0.0332
η_5 educ2	-0.398**	0.0294
η_6 educ3	-0.333**	0.0262
η_7 educ4	-0.133**	0.0309

Equation (6.14)is the employment equation: $y_1^* > 0$ for employed individuals and $y_1^* \leq 0$ for unemployed individuals. Equation (6.15) is the search equation with $y_2^* > 0$ for searchers and $y_2^* \leq 0$ for non-searchers. The variances of the error terms of the employment and the search equation have been normalized to one. The vectors x, z_1 and z_2 contain exogenous individual characteristics. Included in the vector x are log(age) and its square, as well as four education dummies, educ1, educ2, educ3 and educ4, with educ1 the lowest level of education. The highest level of education has been excluded. Included in the vectors z_1 and z_2 are the sum of non-labour income and benefit income, the logarithm of family size, log of age and log of age squared, the four education dummies, three sectoral dummies, sec1, sec2 and sec3, the regional dummies region1, region2 and region3, and a dummy for marital status which is one if married and zero if not. Sec1 is a dummy for education in the technical sector which includes chemistry, physics, mathematics and biology, sec2 refers to economic and administrative education, sec3 is general education and the fourth sector, which serves as reference sector and is not included as a dummy, is the service sector. Region1 is a dummy for the strongly industrialized western part of the Netherlands, Region2 is the east in which there is a mixture of industry and agriculture, Region3 is the south of the Netherlands with some larger companies and agricultural industry and the fourth region, which is the region of reference for which no dummy variable is included, is the remaining part with a sizeable agricultural sector.
Table 6.3 presents the parameter estimates of the wage equation. The double asterisks indicate that the parameters are significant at the 5% level. A single asterisk indicates significance at the 10% level. The age earnings profile reaches its maximum at the age of 44. There is a nice increasing pattern in the parameter estimates of the education dummies, i.e. the higher the level of education, the higher the wage income. Table 6.4 presents the parameter estimates of the employment equation (6.14). Parameter α_2 of non-labour income is significantly negative. Being married and living in the western part of the Netherlands both have a significantly positive effect on the probability of being employed. Educ1, the education dummy associated with the lowest level of education, is significantly negative, wheras educ2, the lowest but one level of education, is significantly negative at the 10% level. Log family size enters positively and significantly at the 10% level. Table 6.5 shows the parameter estimates of the search equation. Both age parameters, β_4 and β_{17} are significant and the probability of searching is maximal at the age of 25, that is, search rises with age until the age of 25, after which it falls. The dummy variables educ3 and region3 are negatively related to search and the relation

Table 6.4 The employment equation

	estimate	standard error
α_1 const	-4.493	3.928
α_2 μ	-0.00168**	0.000366
α_3 log(fs)	0.123*	0.0696
α_4 log(age)	3.354	2.218
α_5 nationality	0.123	0.141
α_6 educ1	-0.665**	0.229
α_7 educ2	-0.363*	0.201
α_8 educ3	-0.112	0.200
α_9 educ4	-0.00995	0.214
α_{10} sec1, technical	-0.0604	0.131
α_{11} sec2, econ. adm.	-0.635	0.157
α_{12} sec3, general	-0.205	0.152
α_{13} marital status	0.650**	0.100
α_{14} region1 (west)	0.260**	0.129
α_{15} region2 (east)	0.0266	0.133
α_{16} region3 (south)	0.204	0.142
α_{17} square of log(age)	-0.501	0.311

Table 6.5 The search equation

	estimate	standard error
β_1 const	-13.322**	3.836
β_2 μ	0.00104**	0.000357
β_3 log(fs)	-0.0530	0.0573
β_4 log(age)	8.265**	2.188
β_5 nationality	-0.110	0.133
β_6 educ1	-0.00303	0.153
β_7 educ2	-0.147	0.125
β_8 educ3	-0.200	0.115
β_9 educ4	0.0256	0.126
β_{10} sec1, technical	-0.129	0.0840
β_{11} sec2, econ. adm.	-0.0718	0.0977
β_{12} sec3, general	-0.0559	0.0975
β_{13} marital status	-0.313**	0.0805
β_{14} region1 (west)	-0.110	0.0980
β_{15} region2 (east)	-0.00105	0.103
β_{16} region3 (south)	-0.209**	0.106
β_{17} square of log(age)	-1.289**	0.312

Table 6.6 The covariances

	estimate	standard error
σ_v (wage)	0.412**	0.00184
σ_{vu_1} (wage-employment)	-0.0482	0.0324
σ_{vu_2} (wage-search)	-0.0260	0.0166
$\sigma_{u_1u_2}$ (employment-search)	-0.858**	0.0204

between being married and search is negative as well. Table 6.6 contains the covariances of the error terms. The wage-employment covariance is not significant. The covariance σ_{vu_2} of wages and the search decision is negative and insignificant. Remarkable is the significance of the age variables in both the wage equation and the search equation. Wages rise with age until the age of 44, whereas search falls with age after the age of 25, so there is a wide range of ages in which wages are rising and search is falling with age. This result is in accordance with the structural model, which implies that there is a negative relation between the search decision and the current wage. Finally, looking at the covariance $\sigma_{u_1u_2}$ between the error terms of employment and search we see that there is a significant and sizeable negative correlation between the search decision and the employment decision, which is in accordance with the large difference in search percentages for employed and unemployed searchers in table 6.1.

In the theoretical model in the preceding section it can be seen from (6.4) that different search equations for the different labour force states arise if there are differences in the arrival rates λ_u and λ_e, the cost functions $c_u(.)$ and $c_e(.)$ or if there are non-zero cost of turnover. With so many sources of possible differences, it is hardly believable that employed and unemployed individuals have the same search equation. Therefore, we repeat the above exercise, but now with different search equations for different labour force states.

Now the model becomes

$$\begin{aligned} \ln w &= \eta' x + v && (6.17)\\ y_1^* &= \alpha' z_1 + u_1 && (6.18)\\ y_e^* &= \beta_e' z_e + u_e && (6.19)\\ y_u^* &= \beta_u' z_u + u_u && (6.20) \end{aligned}$$

The covariance matrix of the disturbances is

$$\Sigma = \begin{pmatrix} \sigma_v^2 & \sigma_{vu_1} & \sigma_{vu_e} & * \\ \sigma_{vu_1} & 1 & \sigma_{u_1u_e} & \sigma_{u_1u_u} \\ \sigma_{vu_e} & \sigma_{u_1u_e} & 1 & * \\ * & \sigma_{u_1u_u} & * & 1 \end{pmatrix} \qquad (6.21)$$

Equation (6.18) is the employment equation. Equation (6.19) is the search equation for employed individuals and equation (6.20) is the search equation for unemployed individuals. Note, that the covariances between the disturbances of search and employment and of search and unemployment are not identified. In the case of cross section data, they do not appear in the likelihood function at all. In table 6.7 the estimates of

Table 6.7 The wage equation

	estimate	standard error
η_1 const	-10.315**	0.902
η_2 log(age)	8.984**	0.507
η_3 square of log(age)	-1.183**	0.0710
η_4 educ1	-0.476**	0.0334
η_5 educ2	-0.400**	0.0294
η_6 educ3	-0.333**	0.0262
η_7 educ4	-0.134**	0.0310

Table 6.8 The employment equation

	estimate	standard error
α_1 const	-5.824	4.082
α_2 μ	-0.00173**	0.000347
α_3 log(fs)	0.118	0.0755
α_4 log(age)	4.133*	2.301
α_5 nationality	0.131	0.144
α_6 educ1	-0.645**	0.229
α_7 educ2	-0.320	0.199
α_8 educ3	-0.102	0.197
α_9 educ4	0.0484	0.211
α_{10} sec1, technical	-0.0490	0.138
α_{11} sec2, econ. adm.	0.0203	0.166
α_{12} sec3, general	-0.198	0.155
α_{13} marital status	0.649**	0.106
α_{14} region1 (west)	0.295**	0.135
α_{15} region2 (east)	0.0397	0.137
α_{16} region3 (south)	0.185	0.148
α_{17} square of log(age)	-0.617*	0.323

Table 6.9 The search equation, Employed individuals

	estimate	standard error
β_{e1} const	-16.440**	5.699
β_{e2} μ	0.000681	0.000554
β_{e3} log(fs)	-0.0531	0.0618
β_{e4} log(age)	10.098**	3.308
β_{e5} nationality	-0.0950	0.149
β_{e6} educ1	-0.0187	0.183
β_{e7} educ2	-0.105	0.133
β_{e8} educ3	-0.195	0.120
β_{e9} educ4	0.0567	0.129
β_{e10} sec1, technical	-0.129	0.0856
β_{e11} sec2, econ. adm.	-0.0658	0.100
β_{e12} sec3, general	-0.0528	0.102
β_{e13} marital status	-0.325**	0.0915
β_{e14} region1 (west)	-0.0895	0.105
β_{e15} region2 (east)	-0.00187	0.108
β_{e16} region3 (south)	-0.229**	0.112
β_{e17} square of log(age)	-1.558**	0.482

Table 6.10 The search equation, Unemployed individuals

	estimate	standard error
β_{u1} const	-12.448	54.753
β_{u2} μ	0.00166	0.00748
β_{u3} log(fs)	1.384	1.041
β_{u4} log(age)	8.948	42.156
β_{u5} nationality	-0.0518	0.784
β_{u6} educ1	-0.264	4.587
β_{u7} educ2	-0.723	3.556
β_{u8} educ3	-0.199	1.292
β_{u9} sec1, technical	0.313	1.159
β_{u10} sec2, econ. adm.	-0.340	1.024
β_{u11} sec3, general	-0.541	2.368
β_{u12} marital status	0.0438	3.895
β_{u13} region1 (west)	-0.397	1.169
β_{u14} region2 (east)	-0.192	0.451
β_{u15} region3 (south)	-0.0940	1.036
β_{u16} square of log(age)	-1.426	6.434

Table 6.11 The covariances

	estimate	standard error
σ_v (wage)	0.411**	0.00194
σ_{vu_1} (wage-employment)	-0.0458	0.0418
σ_{vu_e} (wage-search)	-0.0260	0.0172
$\sigma_{u_1u_u}$ (employment-usearch)	-0.304	6.608
$\sigma_{u_1u_e}$ (employment-esearch)	-0.883**	0.129

the wage equation are given. There is hardly any difference with the estimates in table 6.3. In table 6.8 the estimates of the employment equation are given. The parameter α_2 of μ is significantly negative at the 5% level. In table 6.9 are the estimates of the search equation for employed individuals. There are some differences as compared to the estimates in table 6.5. Non-labour income is no longer significant. The age variables still play an important role and the relation between age and search is maximal at the age of 26. As before, the dumy variables for marital status and region3 have a significant negative effect on the probability of searching while being employed. Table 6.10 presents the search parameter estimates for the unemployed individuals. There are clearly differences between the results in this table and those in 6.5 and 6.9. None of the variables is significant. The large standard errors are due to the low number of observations on non-searching unemployed individuals. Because of the low number of observations we combined the higher two levels of education into one class and consequently the dummy variable educ4 has disappeared from the equation. For this reason, the first model is not nested in the second and therefore a likelihood ratio test for the hypothesis that there is no difference between the two specifications is not valid. Nervertheless, it may be useful to look at the likelihood ratio test statistic to obtain an indication of the difference between the models. The value of the likelihood ratio test statistic 18.9. There are 17 restrictions (17 in the parameters of the search equation, 1 in the covariance structure and -1 because of the combination of two education level dummy variables in the search equation for the unemployed) and the critical value at the 5% level is 27.6, which implies that the hypothesis of no difference between the two models is not rejected. This is mainly due to the limited explanatory power of the search equation for the unemployed individuals. The estimates of the parameters of search equation in first model, in which there is no difference between employed and unemployed individuals, seem to be largely determined by the observations on employed individuals. Finally, table 6.11 shows the parameter estimates of the covariances. The covariance between the wage and employment error terms is insignificant, whereas the correlation between wages and search again is insignificantly negative.

6.3.2 Estimating a structural model of duration and search intensity

It is well-known that there is a close link between labour market transitions and the duration of labour market states. In the case of a stationary model, the duration of being in a certain labour market state is exponentially distributed and its parameter can be derived from the search model as the product of the job offer arrival rate and the job acceptance probability. In the present case we have a two state model in which three types of transitions may occur: Transitions from unemployment into employment with transition rate $\theta_{ue}(s_u)$, job to job transitions with transition rate $\theta_{ee}(s_e)$, and transitions from employment into unemployment with transition rate σ, which is just the layoff rate from section 6.2. Ideally, the arguments s_u and s_e are the observed search intensity variables for unemployed and employed individuals respectively. Unfortunately, search intensity is not an observed variable. In the preceding subsection, we defined four observed indicators $\hat{s}_1, \hat{s}_2, \hat{s}_3$ and $\hat{s}_4$ for the intensity of search. In the present section, the unobserved, latent, search intensity variables s_u and s_e are allowed to be four dimensional vectors of (unobserved) search variables, i.e. $s_u = (s_{u1}, s_{u2}, s_{u3}, s_{u4})'$ and $s_e = (s_{e1}, s_{e2}, s_{e3}, s_{e4})'$, in which each component corresponds to one of the observed indicators. Later on, a relation between the observed instruments $\hat{s}_j$ and the latent variables will be specified.

First the distribution of unemployment duration t_u and job tenure t_e are given for the case of a flow sampling scheme. The density of the latent search intensity vector while employed, conditional on the wage, and the density of the latent search intensity vector while unemployed are denoted by $f_e(s_e)$ and $f_u(s_u)$ respectively. Explicit expressions for those will be given later on. Note, that in the first subsection estimates for the parameters of the wage offer distribution have already been obtained, while accounting for correlation with the employment decision and the search decision. As a consequence, we will condition on the observed wages throughout.

The relation between the transition intensities $\theta_{ue}(s_u)$ and $\theta_{ee}(s_e)$ and the search model is:

$$\begin{aligned} \theta_{ue}(s_u) &= (1+\alpha_u' s_u)\lambda_u \bar{F}(\xi) && (6.22) \\ \theta_{ee}(s_e) &= (1+\alpha_e' s_e)\lambda_e \bar{F}(\alpha(w)) && (6.23) \end{aligned}$$

with $\alpha_u' = (\alpha_{u1}, \alpha_{u2}, \alpha_{u3}, \alpha_{u4})$ and $\alpha_e' = (\alpha_{e1}, \alpha_{e2}, \alpha_{e3}, \alpha_{e4})$. Note, that the parameters α_u and α_e, that determine the influence of search intensity on the transition intensities, can be identified by the duration data.

Now, the density functions of unemployment duration t_u and job tenure t_e, conditional on search intensity s_u and s_e and the wage rate w are:

$$\begin{aligned} f_u(t_u|s_u) &= \theta_{ue}(s_u)\exp\{-\theta_{ue}(s_u)t_u\}, 0 < t_u < \infty \\ f_e(t_e|s_e) &= \theta_{ee}(s_e)\exp\{-(\theta_{ee}(s_e)+\sigma)t_e\}, 0 < t_e < \infty \\ & \qquad \text{job to job transitions} \qquad (6.24) \\ f_e(t_e|s_e) &= \sigma\exp\{-(\theta_{ee}(s_e)+\sigma)t_e\}, 0 < t_e < \infty \\ & \qquad \text{employment to unemployment transitions} \end{aligned}$$

The joint density of duration and the latent search intensity vector can be obtained by multiplying by the marginal density of search intensity, $f_u(s_u)$, for unemployed individuals, or by the density of search intensity conditional on the wage for employed individuals, which we shall simply denote by $f_e(s_e)$.

Before we can derive the likelihood contributions, the relation between the latent search intensity variables and the indicators has to be explained. For this purpose, cost of search functions $c_u(s)$ and $c_e(s)$ are specified for unemployed and employed persons respectively, where $s = (s_1, s_2, s_3, s_4)$, with

$$\begin{aligned} c_u(s) &= \textstyle\sum_{j=1}^4 c_{uj}(s_j) \\ c_{uj}(s_j) &= \gamma_{0u,j} \exp\left(\frac{s_j - \gamma'_{uj}q}{\gamma_{0u,j}}\right) - \gamma_{0u,j} \exp\left(-\frac{\gamma'_{uj}q}{\gamma_{0u,j}}\right) \end{aligned} \tag{6.25}$$

$$\begin{aligned} c_e(s) &= \textstyle\sum_{j=1}^4 c_{ej}(s_j) \\ c_{ej}(s_j) &= \gamma_{0e,j} \exp\left(\frac{s_j - \gamma'_{ej}q}{\gamma_{0e,j}}\right) - \gamma_{0e,j} \exp\left(-\frac{\gamma'_{ej}q}{\gamma_{0e,j}}\right) \end{aligned} \tag{6.26}$$

in which $\gamma_{0u,j}$, $\gamma_{0e,j}$, γ_{uj} and γ_{ej} are parameters and q is a vector of individual characteristics. Note, that at a zero search intensity vector the cost of search is zero. The regularity conditions (6.2) on first and second order derivatives are now assumed to hold for the partial derivatives of the cost function. The regularity conditions for the second order derivatives of the cost of search functions are satisfied if $\gamma_{0u,j}$ and $\gamma_{0e,j}$ are positive, $j = 1, ..., 4$. The positivity conditions on $\gamma_{0e,j}$ and $\gamma_{0u,j}$ will not be imposed in the estimation.

The marginal cost = marginal returns condition (6.4) now becomes

$$\begin{aligned} c'_{uj}(\bar{s}_{uj}) &= R_{uj} \\ c'_{ej}(\bar{s}_{ej}(w)) &= R_{ej}(w), j = 1, 2, 3, 4 \end{aligned} \tag{6.27}$$

with

$$\begin{aligned} R_{uj} &= \alpha_{uj}\lambda_u \textstyle\int_\xi^\infty [W(x) - V] dF(x) \\ R_{ej}(w) &= \alpha_{ej}\lambda_e \textstyle\int_{\alpha(w)}^\infty [W(x) - W(\alpha(w))] dF(x), j = 1, 2, 3, 4 \end{aligned} \tag{6.28}$$

Solving equations (6.27) and (6.28) for each j, using the specification of the cost of search functions in (6.25) and (6.26), we obtain

$$\bar{s}_{uj} = \gamma'_{uj}q + \gamma_{0u,j} \ln R_{uj} = \gamma'_{uj}q + \gamma_{0u,j}[\beta_{uj} + \ln g(\lambda_u, \xi)] \tag{6.29}$$

$$\bar{s}_{ej}(w) = \gamma'_{ej}q + \gamma_{0e,j} \ln R_{ej}(w) = \gamma'_{ej}q + \gamma_{0,e}[\beta_{ej} + \ln g(\lambda_e, \alpha(w))] \tag{6.30}$$

with

$$g(\lambda, w) = \lambda \int_w^\infty [W(x) - W(w)]\, dF(x) \tag{6.31}$$

$$\beta_{uj} = \ln(\alpha_{uj}) \tag{6.32}$$

$$\beta_{ej} = \ln(\alpha_{ej}) \tag{6.33}$$

Note, that $\gamma_{0u,j}$ and $\gamma_{0e,j}$ determine the effect of the marginal returns of search on search intensity. According to the theory specified in section 6.2, the relation between optimal search intensity and the marginal returns of search should be positive. In the empirical application we can test for the sign of this relation.

No explicit expression for the integrand on the right hand side of (6.31) exists and therefore, we need an approximation. The integrand is approximated by [3]

$$(\rho+\sigma)\left[W(x)-W(w)\right] \approx u(x+\mu)-u(w+\mu) \tag{6.34}$$

The function $g(\lambda, w)$ is approximated by $\tilde{g}(\lambda, w)$, with

$$\tilde{g}(\lambda, w)=\frac{\lambda}{\rho+\sigma}\int_w^{\infty}\left[u(x+\mu)-u(w+\mu)\right]dF(x) \tag{6.35}$$

The advantage of this approximation over e.g. a Taylor expansion is that this function has the same derivative properties as the original, i.e. a higher wage leads to lower returns.

Now assume that $\tilde{s}_{uj}$ and $\tilde{s}_{ej}(w)$ are the solutions to

$$\begin{array}{rcl} c'_{uj}(\tilde{s}_{uj}) & = & \alpha_{uj}\tilde{g}(\lambda_u, \xi) \\ c'_{ej}(\tilde{s}_{ej}(w)) & = & \alpha_{ej}\tilde{g}(\lambda_e, \alpha(w)) \end{array} \tag{6.36}$$

Then the latent search intensities s_{uj} and s_{ej} are assumed to be linked to $\tilde{s}_{uj}$ and $\tilde{s}_{ej}(w)$ in the following way:

$$\begin{array}{rclcrclc} \underline{s}_{uj} & = & \tilde{s}_{uj}+\epsilon_{uj}, & & \underline{s}_{ej} & = & \tilde{s}_{ej}(w)+\epsilon_{ej} & \\ s_{uj} & = & \underline{s}_{uj} & \text{if } \underline{s}_{uj}>0, & s_{ej} & = & \underline{s}_{ej} & \text{if } \underline{s}_{ej}>0 \\ s_{uj} & = & 0 & \text{if } \underline{s}_{uj}\leq 0, & s_{ej} & = & 0 & \text{if } \underline{s}_{ej}\leq 0 \end{array} \tag{6.37}$$

with $\epsilon_u=(\epsilon_{u1}, \epsilon_{u2}, \epsilon_{u3}, \epsilon_{u4})' \sim N(0, \Sigma_u)$ and $\epsilon_e=(\epsilon_{e1}, \epsilon_{e2}, \epsilon_{e3}, \epsilon_{e4})' \sim N(0, \Sigma_e)$. The relation with the observed indicator becomes:

$$\begin{array}{rcllcl} \hat{s}_{uj} & = 1 & \text{if} & s_{uj} > 0, & \hat{s}_{u4} = s_{u4} & \text{if } s_{u4} > 0 \\ & = 0 & \text{if} & s_{uj} = 0, j=1,2,3, & \quad = 0 & \text{if } s_{u4} = 0 \end{array} \tag{6.38}$$

For non-searchers, $\hat{s}_{uj}=0, j=1,2,3,4$. The same holds for employed individuals. The variances of $\epsilon_{zj}, z=e,u, j=1,2,3$, are normalized to one.

For all of the four search indicators we observe whether or not it is positive, but if it is positive, its value is only observed for $\hat{s}_4$, the number of applications. Therefore we have to integrate out the unobserved values. Let $f_u^*(\underline{s}_u)$ denote the density function of $\underline{s}_u$, which is normal according to (6.37) and $f_e^*(\underline{s}_e)$ denotes the density of $\underline{s}_e$. Then the joint density of unemployment duration and the indicator vector becomes:

$$\begin{array}{l} f_u(t_u, \hat{s}_u) = \\ \int\int_{A(\hat{s})}\int \theta_{ue}(I'_u\underline{s}_u)\exp\{-\theta_{ue}(I'_u\underline{s}_u)t_u\} f_u^*(\hat{s}_4|\underline{s}_{u1}, \underline{s}_{u2}, \underline{s}_{u3}) f_u^*(\underline{s}_{u1}, \underline{s}_{u2}, \underline{s}_{u3}) d\underline{s}_{u1}d\underline{s}_{u2}d\underline{s}_{u3} \\ \qquad \text{if } \hat{s}_4>0 \\ \qquad = \\ \int\int_{B(\hat{s})}\int\int \theta_{ue}(I'_u\underline{s}_u)\exp\{-\theta_{ue}(I'_u\underline{s}_u)t_u\} f_u^*(\underline{s}_{u1}, \underline{s}_{u2}, \underline{s}_{u3}, \underline{s}_{u4}) d\underline{s}_{u1}d\underline{s}_{u2}d\underline{s}_{u3}d\underline{s}_{u4} \\ \qquad \text{if } \hat{s}_4=0 \end{array} \tag{6.39}$$

[3] Note that the integrand can be written as $(\rho+\sigma)\left[W(x)-W(w)\right] = u(x+\mu)-u(w+\mu)-[c_e(s_e^*(x))-c_e(s_e^*(w))]+(1+\alpha_e s_e^*(x))\lambda_e\int_{\alpha(x)}^{\infty}[W(\tilde{x})-W(\alpha(x))]dF(\tilde{x})-(1+\alpha_e s_e^*(w))\lambda_e\int_{\alpha(w)}^{\infty}[W(\tilde{x})-W(\alpha(w))]dF(\tilde{x})$ If $s_e^*(x)-s_e^*(w)$ is small, but not equal to zero, this expression can be approximated by: $(\rho+\sigma)\left[W(x)-W(w)\right]\approx u(x+\mu)-u(w+\mu)+(s_e^*(x)-s_e^*(w))\left\{-c'_e(s_e^*(w))+R_e(w)\right\}$ The term in brackets is the difference between the marginal cost of search and the marginal returns of search, which, in the case of non-zero optimal search intensity, is zero.

$A(\hat{s})$ and $B(\hat{s})$ are three and four-dimensional regions of integration, respectively, which depend on the value of the indicator $\hat{s}$. The bounds of integration are $(-\infty, 0)$ if $\hat{s}_j = 0$ and $(0, \infty)$ if $\hat{s}_j = 1$. I_u is a four-dimensional indicator function, with $I_{uj} = 0$ if $\underline{s}_{uj} \leq 0$ and $I_{uj} = 1$ if $\underline{s}_{uj} > 0$, indicating that if there is no search, there is no effect on the job offer arrival rate. For employed individuals the density becomes

$$\begin{array}{l}
f_e(t_e, \hat{s}_e) = \\
\int \int_{A(\hat{s})} \int \theta_{ee}(I_e' \underline{s}_e) \exp\{-(\theta_{ee}(I_e' \underline{s}_e) + \sigma) t_e\} f_e^*(\hat{s}_4 | \underline{s}_{e1}, \underline{s}_{e2}, \underline{s}_{e3}) f_e^*(\underline{s}_{e1}, \underline{s}_{e2}, \underline{s}_{e3}) d\underline{s}_{e1} d\underline{s}_{e2} d\underline{s}_{e3} \\
\qquad \text{if } \hat{s}_4 > 0 \\
\int \int_{B(\hat{s})} \int \int \theta_{ee}(I_e' \underline{s}_e) \exp\{-(\theta_{ee}(I_e' \underline{s}_e) + \sigma) t_e\} f_u^*(\underline{s}_{e1}, \underline{s}_{e2}, \underline{s}_{e3}, \underline{s}_{e4}) d\underline{s}_{e1} d\underline{s}_{e2} d\underline{s}_{e3} d\underline{s}_{e4} \\
\qquad \text{if } \hat{s}_4 = 0 \\
\qquad \text{(job to job transitions)} \\
\quad = \\
\int \int_{A(\hat{s})} \int \sigma \exp\{-(\theta_{ee}(I_e' \underline{s}_e) + \sigma) t_e\} f_e^*(\hat{s}_4 | \underline{s}_{e1}, \underline{s}_{e2}, \underline{s}_{e3}) f^*(\underline{s}_{e1}, \underline{s}_{e2}, \underline{s}_{e3}) d\underline{s}_{e1} d\underline{s}_{e2} d\underline{s}_{e3} \\
\qquad \text{if } \hat{s}_4 > 0 \\
\int \int_{B(\hat{s})} \int \int \sigma \exp\{-(\theta_{ee}(I_e' \underline{s}_e) + \sigma) t_e\} f_u^*(\underline{s}_{e1}, \underline{s}_{e2}, \underline{s}_{e3}, \underline{s}_{e4}) d\underline{s}_{e1} d\underline{s}_{e2} d\underline{s}_{e3} d\underline{s}_{e4} \\
\qquad \text{if } \hat{s}_4 = 0 \\
\qquad \text{(employment to unemployment transitions)}
\end{array} \tag{6.40}$$

I_e is a three-dimensional indicator function with $I_{ej} = 0$ if $\underline{s}_{ej} \leq 0$ and $I_{ej} = 1$ if $\underline{s}_{ej} > 0$.

To evaluate the likelihood contribution, we need to calculate three and four-dimensional integrals of normally distributed random variables. This problem can easily be handled by using the smooth recursive conditioning algorithm (SRC) for simulating multidimensional integrals over normally distributed random variables and applying simulated maximum likelihood (SML) as described in Börsch-Supan and Hajivassiliou (1993).

The endogeneity of search intensity has consequences for the joint density in the case of a stock sampling scheme. The derivation of the joint density of duration and search intensity, conditional on backward recurrence times is given in appendix B.

The reservation wage for unemployed individuals can be calculated by solving ξ from the implicit equation (6.8). The solution can be used to determine search intensity equation. To calculate the transition intensity $\theta_{ue}(s_u)$, s_u^* is replaced by s_u in (6.8).

For employed individuals, the reservation wage $\alpha(w)$ is calculated by means of the Taylor approximation (6.11), where $s_e^{**}(w)$ is replaced by s_e in the calculation of the conditional distribution of duration .

In the estimation of the structural model the following approach will be followed. First of all the parameters of the wage distribution will be fixed to the parameters obtained by the estimation of the reduced form model, described in the previous subsection. From the structural model in section 6.2 it has become clear that a structural specification of the distribution of observed wages would depend, in a complicated way, on the employment situation and the search decision. For a structural model of the wage parameters, it is not enough to know the labour market state and the income variable at the point of sampling: Information about previous labour market states is also required. Although the use of reduced form estimators of the wage parameters will lead to inefficient estimators of the remaining model parameters, in practice the gains of estimating the wage parameters structurally will not outweigh the cost that is due to the intractibility of the model.

A joint model of duration and search intensity will be estimated, conditional on the wage. The parameters α_u, α_e, λ_u, λ_e, γ_u, γ_e, $\gamma_{0,u}$, $\gamma_{0,e}$, ω, σ and k can be obtained by the (simulated) maximum likelihood principle.

For the utility function $u(x)$ a linear specifications is chosen. According to (6.1) the linear specification is:

$$\begin{aligned} \text{utility(income=x,state=employed)} &= x \\ \text{utility(income=x,state=unemployed)} &= \omega x \end{aligned} \tag{6.41}$$

The cost of turnover can be parametrized as:

$$k = \delta' c \tag{6.42}$$

in which δ is a parameter vector and c is a vector of individual characteristics. λ_u and λ_e can be parametrized by

$$\begin{aligned} \lambda_u &= \exp(\kappa_u' z) \\ \lambda_e &= \exp(\kappa_e' z) \end{aligned} \tag{6.43}$$

in which z is a vector of individual characteristics and κ_u and κ_e are parameter vectors.

To restrict the number of parameters we assume that

$$\begin{aligned} \gamma_{uj} &= \vartheta_{uj}\gamma_u, \vartheta_{u1} = 1, j = 1,2,3,4 \\ \gamma_{ej} &= \vartheta_{ej}\gamma_e, \vartheta_{e1} = 1, j = 1,2,3,4 \end{aligned} \tag{6.44}$$

where ϑ_{uj} and ϑ_{ej} are scalars. The economic argument in favour of this type of restriction is that the four measures $\hat{s}_j$ are all indicators for search effort. The restriction implies that there is a single index, $\gamma_u' q$ for unemployed and $\gamma_e' q$ for employed individuals, which specifies the effect of individual characteristics on cost of search. Suppose, that there is a variable called "search effort", which is a weighted sum of the latent variables s_{uj} and s_{ej}, for unemployed and employed individuals respectively. Then the structural equation for this variable would be of the same form as (6.29)-(6.30)-(6.37), i.e. linear in q and and log-linear in R_u or R_e, and the cost of search function of this variable would be of the same form as $c_{uj}(.)$ and $c_{ej}(.)$ in (6.25) and (6.26).

6.3.3 Data and estimation results

Table 6.12 provides sample statistics of the sample of employment search spells as well as the sample of unemployment search spells. Table 6.13 presents sample statistics for spells of non-searchers. In the sample we only consider single spells of employment and unemployment. This is done for several reasons. First of all, to relax the effects of the stationarity assumption. If we would estimate a multiple spell model, we would have to assume that the individual characteristics in the transitions intensities are the same throughout all spells of employment and unemployment, which is rather unrealistic. Second, the indicators of search intensity are only observed twice a year, in the months in which the survey is conducted. Therefore, for shorter spells which are in between the survey months, search intensity is not observed. Third, using a multiple spell model it would not be possible anymore to condition on the wage rate, which makes the model

Table 6.12 Sample statistics, searchers

Employed		
variable	mean	standard deviation
age	32	7.8
family size (persons)	3.2	1.3
education level	mode 3	
Dutch nationality	96.6%	
region 1 (industrialized west)	44.4%	
region 2 (east)	24.8%	
region 3 (south)	20.8%	
region 4 (agricultural)	10.0%	
married	65.6%	
sector of education 1 (technical)	28.6%	
sector of education 2 (economic/administrative)	17.4%	
sector of education 3 (no specialization)	27.6%	
sector of education 4 (services)	23.4%	
Unemployed		
variable	mean	standard deviation
age	31	11.7
family size (persons)	3.0	1.6
education level	mode 1	
Dutch nationality	92.3%	
region 1 (industrialized west)	37.2%	
region 2 (east)	27.9%	
region 3 (south)	23.4%	
region 4 (agricultural)	11.5%	
married	35.5%	
sector of education 1 (technical)	24.4%	
sector of education 2 (economic/administrative)	10.6%	
sector of education 3 (no specialization)	52.6%	
sector of education 4 (services)	11.2%	

Table 6.13 Sample statistics, non-searchers

Employed		
variable	mean	standard deviation
age	38	10.7
family size (persons)	3.3	1.3
education level	mode 3	
Dutch nationality	96.4%	
region 1 (industrialized west)	43.0%	
region 2 (east)	23.8%	
region 3 (south)	23.1%	
region 4 (agricultural)	10.1%	
married	76.6%	
sector of education 1 (technical)	34.4%	
sector of education 2 (economic/administrative)	17.8%	
sector of education 3 (no specialization)	27.7%	
sector of education 4 (services)	19.3%	
Unemployed		
variable	mean	standard deviation
age	47	15.4
family size (persons)	2.4	1.2
education level	mode 1	
Dutch nationality	95.0%	
region 1 (industrialized west)	42.5%	
region 2 (east)	35.0%	
region 3 (south)	5.0%	
region 4 (agricultural)	17.5%	
married	57.5%	
sector of education 1 (technical)	17.5%	
sector of education 2 (economic/administrative)	10.0%	
sector of education 3 (no specialization)	50.0%	
sector of education 4 (services)	20.0%	

intractable. The following sampling scheme is used. All individuals who are either employed or unemployed, are sampled from the survey wave of October 1987. In addition, all individuals who are employed or unemployed in April 1988, and whose spell started after October 1987, have been sampled from the wave of April 1988. Finally, all spells of individuals who are unemployed in October 1988, and whose spell started after April 1988 have been added. In all cases we have to correct for the fact that we do not observe spells in between the survey months, which can be done by using the stock sample density for all observations, which is derived in appendix B.

The total sample of spells of employed searchers consists of 500 observations. 127 of these spells are completed spells. The observation period ends in October 1988. The sample of spells of unemployed searchers consists of 312 observations. 139 of these spells are completed spells. For the spells of unemployed individuals, the observation period ends in April 1989.

The sample of spells of non-searchers consists of 2806 observations, which can be divided into 2766 employment spells and 40 unemployment spells. For the employed, 171 job to job transitions occur, whereas 165 transitions into unemployment take place. The number of unemployment spells that is completed is 13.

Comparing spells of searchers and spells of non-searchers we can say that 33% of the non-search unemployment spells ends with a transition into employment, whereas the number is 45% for search unemployment spells. For the non-search employment spells 6% ends with a job to job transition, whereas 25% of the search employment spells ends with a job to job transition. 6% of the non-search employment spells ends by a transition into unemployment whereas the number is 4% for the search employment spells. In the vector z in (6.43) the following characteristics are included: A constant term, log of age and log of age squared, the three sectoral dummies sec1, sec2, and sec3, that have already been described in section 6.3.1, the regional dummy variables region1, region2 and region3, the three education dummies educ1, educ2 and educ3, a dummy for marital status and a dummy for nationality. There are too few observations to make the layoff rate σ dependent on individual characteristics. In the vector q, which contains characteristics of the cost of search functions in (6.25) and (6.26), we include the logarithm of family size, a dummy variable for marital status, as well as the logarithm of age and its square. In the turnover cost k in (6.42) we include two age dummy variables: agedum1 for individuals aged 30 or lower, and agedum2 for individuals aged between 30 and 45. Two dummy variables are included which are concerned with the training period required for the worker's present job. If an individual had to spend much time to get settled in his present job, turnover cost is likely to be higher than in absence of a training period. The first dummy variable is called trainper1, which is equal to one if there is no training period or if the respondent does not know the length of the training period. The second dummy, trainper2, is for persons with a training period with a length that is less than six months. The reference group includes those individuals who have a job for which a training period is required that is half a year or longer.

The estimation results of the structural model are reported in the tables 6.14 through 6.20. The rate of time preference ρ has been fixed, such that on a yearly basis the discount rate is 5%. The number of replications used for simulating probability integrals by means of the smooth recursive conditioning algorithm is 20. In the table a double asterisk indicates significance at the 5% level, whereas a single asterisk indicates significance at the 10% level.

First we discuss the parameters λ_e which establish the part of the job offer arrival rate for employed individuals that is determined by the demand side of the labour market. The estimates are presented in table 6.14. The dummy variable sec2 (parameter κ_{e5}), which corresponds with the economic and administrative sector, has a significantly positive effect on λ_e. As the reference sector is the sector of services, this implies that individuals with an economic or administrative education have a significantly higher job offer arrival rate than individuals in the service sector. The education dummy educ1 is significant. Both age variables, log(age) and its square, are significant determinants of the arrival rate. The values of the estimates indicate that λ_e falls with age at an increasing rate.

The parameter estimate of the layoff rate σ is small but significant.

Comparing the parameter estimates of the effectiveness of search parameters α_{e1}, α_{e2} and α_{e4}, which determine the effectiveness of search of respectively attitude, screening and number of applications, we see that attitude has a large effect on the probability of getting a job offer as compared to the remaining two search instruments. In comparing the effectiveness of the three search instruments, the scale of the variables has to be taken into account, i.e. the variances of the attitude and screening variables are normalized to one, whereas the variance of the application instrument has been estimated. Even after accounting for scale, the effect of the number of applications on the number of job offers remains surprisingly low. The standard errors of $\hat{\alpha}_{e2}$ (screening) and $\hat{\alpha}_{e4}$ (applications) are high relative to the parameter estimates. Note that a formal test for $\alpha_{ej} = 0$ is not possible: Under the null, marginal returns of search would always be zero and consequently nobody would search. The fact that individuals <u>are</u> making use of the search instruments indicates that at least they <u>expect</u> search to be effective.

The estimates of the cost of search function parameters for employed individuals, $\gamma_{ej}, j = 1, ..., 5$ are presented in table 6.15. The characteristics that are related to the household composition do not play any role. The age variables (parameters γ_{e4} and γ_{e5}) are significant. The cost of search falls with age until the age of 30, after which it rises. The parameters $\gamma_{0e,j}$ which determine the effect of marginal returns of search on the intensity of search are all positive and significant, which is in accordance with the regularity conditions that are assumed to hold in the economic model.

The estimates of the cost of turnover parameters, $\delta_1, ..., \delta_5$, are presented in table 6.16. Note, that the cost of turnover, as it has been introduced here, only serves as a relaxation of the functional form of the transition intensity $\theta_{ee}(s_e)$, as no data on reservation wages are available.[4] This means that we have to be careful in interpreting the estimates of turnover costs. None of the cost of turnover parameters are signifiantly different from zero.

The arrival rate parameters λ_u for unemployed individuals can be found in table 6.18. The age parameters are significant. The estimates suggest that λ_u rises with age, which seems counter-intuitive. Living in the southern part of the Netherlands (region 3, parameter κ_{u9}) adds positively to the job offer arrival rate. The parameter estimate of

[4]Van den Berg (1992) uses data on reservation wages to estimate the turnover costs by minimizing the squared difference between observed reservation wages and their theoretical expressions.

the education dummy variable educ2 is significantly positive.

The utility parameter ω exceeds one, but is not significantly different from one, which means that we cannot say that there are differences in utility levels associated with the two different labour market states.

As for the employed individuals, we see that the attitude variable influences the arrival rate strongly. For unemployed individuals the number of applications has much more effect than for the employed individuals. Being registered at the employment office seems to be the least effective instrument of search. Comparing the parameter estimates with the standard errors, we see that the standard errors for attitude, screening and applications are low as compared to the the estimates. For the employment office parameter this is not the case.

The estimates of the cost of search parameters $\gamma_{uj}, j = 1, ..., 5$, are shown in table 6.19. Family size has a significant and positive effect on the intensity of search. The larger the family, the more effort is spent on search. In the extended search model of Burdett and Mortensen (1978) cost of search arises from a loss in utility, as a higher intensity of search implies that less time can be devoted to leisure. From this point of view, having a larger family affects preferences such that the individual is willing to allocate less time to leisure and more time to search. Cost of search is increasing with age. The parameters $\gamma_{0u,j}$, that determine the effect of marginal returns of search on search intensity, all are significantly positive, which again is in accordance with the assumptions underlying the economic model.

In conclusion we can say that, for both unemployed searchers and employed searchers, the theoretical result that marginal returns of search should increase the intensity of search, is not rejected by the estimation results. For employed individuals, search intensity varies with age: it increases with age until the age of 30, after which it decreases. For unemployed persons, family size is an important determinant of search intensity. If we consider the measure of search intensity to be a proxy for time spent on search, the significance of family size can be explained in terms of allocation of time to search activities and leisure. For both employed and unemployed individuals, the attitude seems to be an important determinant of search intensity in terms of effectiveness of search. For employed individuals, the number of applications does not seem to be very effective, whereas for unemployed individuals the number of applications is effective. Being registered at the employment office is the least effective instrument of search for the unemployed individuals.

Table 6.14 Estimates of the structural model
Employed individuals
The arrival rates

	estimate	standard error
CHARACTERISTICS OF λ_e		
κ_{e1} (const)	-9.757**	1.558
κ_{e2} (log(age))	4.720**	0.877
κ_{e3} (square of log(age))	-0.967**	0.136
κ_{e4} (sec1)	0.023	0.168
κ_{e5} (sec2)	0.343*	0.187
κ_{e6} (sec3)	-0.005	0.194
κ_{e7} (region1)	0.205	0.198
κ_{e8} (region2)	-0.034	0.220
κ_{e9} (region3)	0.097	0.214
κ_{e10} (educ1)	0.503**	0.238
κ_{e11} (educ2)	0.308*	0.178
κ_{e12} (educ3)	0.031	0.162
κ_{e13} (marital status)	0.055	0.148
κ_{e14} (nationality)	0.329	0.277
THE LAYOFF RATE		
σ	0.0016**	0.00023
EFFECTIVENESS OF SEARCH		
α_{e1} (attitude)	25.739	4.092
α_{e2} (screening)	0.156	0.161
α_{e4} (applications)	0.0022	0.0028

Table 6.15 Estimates of the structural model
Employed individuals
The cost of search function

	estimate	standard error
COST OF SEARCH INDICATOR $\gamma_e' q$		
γ_{e1} (constant)	-21.40**	4.612
γ_{e2} (log(family size))	-0.041	0.060
γ_{e3} (marital status)	0.075	0.080
γ_{e4} (log(age))	11.70**	2.967
γ_{e5} (square of log(age))	-1.722**	0.393
EFFECT OF RETURNS OF SEARCH ON SEARCH INTENSITY		
$\gamma_{oe,1}$ (attitude)	0.087**	0.043
$\gamma_{oe,2}$ (screening)	0.176**	0.044
$\gamma_{oe,4}$ (applications)	0.864**	0.177
PARAMETER ϑ_e		
ϑ_{e2} (screening)	0.849**	0.109
ϑ_{e4} (applications)	1.452**	0.683

Table 6.16 Estimates of the structural model
Employed individuals
Cost of turnover

	estimate	standard error
δ_1 (constant)	-9.641	9.044
δ_2 (agedum1)	190.0	241.3
δ_3 (agedum2)	-109.7	226.0
δ_4 (trainper1)	385.8	285.0
δ_5 (trainper2)	247.2	278.7

Table 6.17 Estimates of the structural model
Employed individuals
Parameters of error distribution, Σ_e

	estimate	standard error
$\sigma_{e,12}$ (covariance attitude-screening)	0.936**	0.009
$\sigma_{e,14}$ (covariance attitude-applications)	1.938**	0.067
$\sigma_{e,24}$ (covariance screening-applications)	2.097**	0.063
$\sigma_{e,4}$ (standard deviation applications)	2.248**	0.072

Table 6.18 Estimates of the structural model
Unemployed individuals
The arrival rates

	estimate	standard error
CHARACTERISTICS OF λ_u		
κ_{u1} (const)	4.336**	1.840
κ_{u2} (log(age))	-2.638**	0.752
κ_{u3} (square of log(age))	0.453**	0.174
κ_{u4} (sec1)	0.341	0.509
κ_{u5} (sec2)	0.581	0.630
κ_{u6} (sec3)	0.284	0.467
κ_{u7} (region1)	-0.291	0.462
κ_{u8} (region2)	-0.278	0.455
κ_{u9} (region3)	1.051**	0.478
κ_{u10} (educ1)	0.635	0.525
κ_{u11} (educ2)	1.199**	0.498
κ_{u12} (educ3)	0.542	0.504
κ_{u13} (marital status)	0.312	0.414
κ_{u14} (nationality)	0.393	0.359
UTILITY PARAMETER		
ω	1.260**	0.270
EFFECTIVENESS OF SEARCH		
α_{u1} (attitude)	23.999	8.232
α_{u2} (screening)	0.341	0.154
α_{u3} (employment office)	0.097	0.056
α_{u4} (applications)	1.507	0.506

Table 6.19 Estimates of the structural model
Unemployed individuals
The cost of search function

	estimate	standard error
COST OF SEARCH INDICATOR $\gamma_u' q$		
γ_{u1} (constant)	-1.906**	0.314
γ_{u2} (log(family size))	0.168**	0.021
γ_{u3} (marital status)	-0.030	0.029
γ_{u4} (log(age))	-0.665**	0.261
γ_{u5} (square of log(age))	0.025	0.039
EFFECT OF RETURNS OF SEARCH ON SEARCH INTENSITY		
$\gamma_{0u,1}$ (attitude)	0.450**	0.024
$\gamma_{0u,2}$ (screening)	1.523**	0.349
$\gamma_{0u,3}$ (employment office)	1.857**	0.349
$\gamma_{0u,4}$ (applications)	1.291**	0.222
PARAMETER ϑ_u		
ϑ_{u2} (screening)	1.478**	0.494
ϑ_{u3} (employment office)	1.366**	0.437
ϑ_{u4} (applications)	1.043**	0.361

Table 6.20 Estimates of the structural model
Unemployed individuals
Parameters of error distribution, Σ_u

	estimate	standard error
$\sigma_{u,12}$ (covariance attitude-screening)	-0.091**	0.017
$\sigma_{u,13}$ (covariance attitude-employment office)	0.183*	0.107
$\sigma_{u,14}$ (covariance attitude-applications)	-2.679**	0.292
$\sigma_{u,23}$ (covariance screening-employment office)	0.292**	0.107
$\sigma_{u,34}$ (covariance employment office-applications)	0.380	0.396
$\sigma_{u,24}$ (covariance screening-applications)	2.400**	0.264
$\sigma_{u,4}$ (standard deviation applications)	4.384**	0.314

6.4 Conclusions

We have specified an empirical version of the search model of Mortensen (1986), in which the intensity of search is a choice variable for the individual. A higher level of search intensity increases the job offer arrival rate, but at the same time cost of search rises. The individual chooses the intensity of search on the basis of a comparison of marginal returns of search with marginal cost of search. We allowed for differences in arrival rates between the state of employment and the state of unemployment. This means that there are differences in search conditions for different labour force states. We have seen that these differences in search conditions affect the reservation wage for individuals in the state of unemployment. The better the search conditions in the state of unemployment, as compared to the search conditions in the state of employment, the higher the reservation wage for individuals in the state of unemployment.

If cost of turnover is zero, the reservation wage in the employment state is equal to the present wage. Positive turnover cost raises the reservation wage in the employment state. As higher cost of turnover deteriorates the search conditions while employed, the reservation wage for unemployed persons rises as well.

According to the assumptions underlying the economic model, higher marginal returns of search should increase the intensity of search. In the empirical application, we have tested for this implication of economic theory and it could not be rejected.

In the empirical model we used data on job duration and unemployment duration and several indicators of search intensity to estimate the model parameters. The cost of search for employed persons is largely determined by age patterns. Cost of search decrease until the age of 30, after which it increases. For unemployed persons family size is a significant determinant of cost of search. A larger family leads to a lower cost of search and consequently to a higher intensity of search. Cost of search is increasing with age for unemployed persons.

We have not found significant evidence in favour of differences in preferences with respect to different labour market states.

Estimates of the arrival rate reveal evidence about the relative effectiveness of different search instruments. For both employed and unemployed persons, the search intensity indicator that measures the individual's attitude towards search strongly influences the arrival rate. This may be partly due to the possibility that the attitude variable picks up search intensity channels that cannot be assigned to one of the remaining three indicators of search. Rather surprisingly, we found that the number of applications has not much effect for employed individuals. For unemployed individuals the number of applications has a stronger effect on the arrival rate. Being registered at the employment office is the least effective search instrument for unemployed persons.

6.A The Bellman equations

First the expression of the value function, V, for unemployed individuals is derived. We consider the events in a small time interval of length Δt. At present, the individual is unemployed, is earning a benefit income of b and a non-labour income μ. The within period utility flow in a time interval with length Δt is $\omega u(b+\mu)\Delta t$. If the individual is searching during Δt, the cost of search is $c_u(s)\Delta t$. The current period contribution to V is $(\omega u(b+\mu) - c_u(s))\Delta t$. At the end of the interval Δt he may or may not obtain a job offer. The number of job offers obtained in an interval of length Δt is Poisson distributed with parameter $(1+\alpha_u s)\lambda_u \Delta t$, so the probability of receiving a job offer is $e^{-(1+\alpha_u s)\lambda_u \Delta t}(1+\alpha_u s)\lambda_u \Delta t + o(\Delta t)$, whereas the probability of receiving no job offer is $1 - e^{-(1+\alpha_u s)\lambda_u \Delta t}(1+\alpha_u s)\lambda_u \Delta t + o(\Delta t)$. If a job offer is obtained with wage income x a choice can be made between accepting the job with value $W(x)$ or rejecting with value V. Therefore, the expected future value is $E_x \max[V, W(x)]$. In the absence of a job offer the value remains equal to V. The discount factor is $e^{-\rho \Delta t}$. The value function becomes:

$$\begin{aligned} V = \; & \max_{s\geq 0} \left[(u(b+\mu) - c_u(s))\, \Delta t + \right. \\ & .e^{-\rho\Delta t} \left\{ e^{-(1+\alpha_u s)\lambda_u \Delta t}(1+\alpha_u s)\lambda_u \Delta t E_x \max[V, W(x)] \right. \\ & \left.\left. +(1 - e^{-(1+\alpha_u s)\lambda_u \Delta t}(1+\alpha_u s)\lambda_u \Delta t) V \right\} \right] + o(\Delta t) \end{aligned} \tag{6.A.1}$$

Rearranging terms yields:

$$\begin{aligned} & \tfrac{1-e^{-\rho\Delta t}}{\Delta t} V = \\ & \max_{s\geq 0} \left[\omega u(b+\mu) - c_u(s) + e^{-(\rho+(1+\alpha_u s)\lambda_u)\Delta t}(1+\alpha_u s)\lambda_u \left\{ E_x \max[V, W(x)] - V \right\} \right] \end{aligned} \tag{6.A.2}$$

Letting $\Delta t \to 0$:

$$\rho V = \max_{s\geq 0} \left[\omega u(b+\mu) - c_u(s) + (1+\alpha_u s)\lambda_u \left\{ E_x \max[V, W(x)] - V \right\} \right] \tag{6.A.3}$$

Replacing the expectation sign by the integral over the wage distribution yields the first equation of (6.3).

For individuals who are currently working at wage w the value function is denoted by $W(w)$. The current period contribution to the value function is $(u(w+\mu) - c_e(s))\Delta t$. At the end of interval Δt four events may occur. With probability $(e^{-(1+\alpha_e s)\lambda_e \Delta t}(1+\alpha_e s)\lambda_e \Delta t + o(\Delta t))(1-\sigma\Delta t) = e^{-(1+\alpha_e s)\lambda_e \Delta t}(1+\alpha_e s)\lambda_e \Delta t + o(\Delta t)$ a job offer with wage x is obtained by the individual, while he is not laid off. If he accepts, he has to pay turnover costs k. The alternatives are, to remain in his present job or to become unemployed. The value for the event is $\max[V, W(x) - k, W(w)]$. The second event is that of getting a job offer and being laid off. Now the alternative of remaining in his present job disappears. The probability of the event is $(e^{-(1+\alpha_e s)\lambda_e \Delta t}(1+\alpha_e s)\lambda_e \Delta t + o(\Delta t))\sigma \Delta t = o(\Delta t)$ with value $\max[V, W(x) - k]$. The third event is that of neither getting a job offer, nor being laid off. The probability is $(1 - e^{-(1+\alpha_e s)\lambda_e \Delta t}(1+\alpha_e s)\lambda_e \Delta t + o(\Delta t))(1 - \sigma\Delta t) = 1 - e^{-(1+\alpha_e s)\lambda_e \Delta t}(1+\alpha_e s)\lambda_e \Delta t - e^{-(1+\alpha_e s)\lambda_e \Delta t}(1+\alpha_e s)\lambda_e \Delta t + o(\Delta t)$ with value $\max[V, W(w)] = W(w)$. Finally, he may be laid off without getting a job offer. This event has probability $(1 - e^{-(1+\alpha_e s)\lambda_e \Delta t}(1+\alpha_e s)\lambda_e \Delta t + o(\Delta t))\sigma\Delta t = \sigma\Delta t + o(\Delta t)$. The value function becomes:

$$
\begin{aligned}
&W(w) = \\
&\max_{s\geq 0}\left[(u(w+\mu)-c_e(s))\Delta t\right. \\
&+e^{-\rho\Delta t}\left\{e^{-(1+\alpha_e s)\lambda_e\Delta t}(1+\alpha_e s)\lambda_e\Delta E_x \max[V, W(x)-k, W(w)]+\right. \\
&\left.\left.(1-e^{-(1+\alpha_e s)\lambda_e\Delta t}(1+\alpha_e s)\lambda_e\Delta t-\sigma\Delta t)W(w)+\sigma\Delta t V\right\}\right]+o(\Delta t)
\end{aligned}
\tag{6.A.4}
$$

Rearranging terms, dividing by Δt and letting $\Delta t \to 0$ yields

$$
\begin{aligned}
&(\rho+\sigma)W(w) = \\
&\max_{s\geq 0}\left[u(w+\mu)-c_e(s)+\lambda_e s\left\{E_x[V, W(x)-k, W(w)]-W(w)\right\}+\sigma V\right]
\end{aligned}
\tag{6.A.5}
$$

This is the equivalent of the second equation in (6.3).

6.B The stock sample density

The stock sample density of duration and search intensity, conditional on the backward recurrence time is derived. The analysis is based on Ridder (1984). The subindices e and u, indicating the labour force state, will be suppressed. Let $f(t|s,w)$ denote the flow conditional density of duration, conditional on search intensity and the wage. To reduce the necessary notation, search intensity is treated as a observed continuous non-negative random variable here. The extension to multidimensional variables of the type in section 6.3.2 is straightforward. Let $f(s|w)$ denote the density of search intensity conditional on the wage, and let $g(w)$ denote the marginal density of observed wages. Then the joint flow density of duration, search intensity and observed wages is

$$f(t|s,w)f(s|w)g(w), 0<t<\infty, 0<s<\infty \tag{6.B.1}$$

Now assume that the inflow rate into the given labour force state is $i(-p,l)$, in which $-p$ denotes the time of inflow into the state, if the point of sampling is taken as reference, and l is calendar time. The stock density is the flow density, conditional on entrance at p time units ago, and conditional on duration t exceeding the backward recurrence time p. Then the joint stock density of duration, backward recurrence time, search intensity and observed wages is: [5]

$$
\begin{aligned}
h(p,t,s,w) = \;&\frac{i(-p,l)f(t|s,w)f(s|w)g(w)}{\int_0^\infty\int_0^\infty\int_0^\infty i(-\tilde{p},l)F(\tilde{p}|\tilde{s},\tilde{w})f(\tilde{s}|\tilde{w})g(\tilde{w})d\tilde{w}d\tilde{s}d\tilde{p}} \\
&0<p<\infty \\
&p<t<\infty \\
&0<s<\infty
\end{aligned}
\tag{6.B.2}
$$

We are interested in the stock density of duration and search intensity, conditional on the wage and the backward recurrence time, i.e.

$$h(t,s|p,w) = \frac{h(p,t,s,w)}{h(p,w)} \tag{6.B.3}$$

[5] Note, that we treat the subsample of employment spells and the subsample of unemployment spells as two separate samples here. Treating them as one sample changes the selectivity correction in $h(p,t,s,w)$, but leaves the final result, i.e. the density conditional on backward recurrence times, unaffected.

in which

$$h(p,w) = \int_0^\infty \int_p^\infty h(p,\tilde{t},\tilde{s},w)d\tilde{t}d\tilde{s} \tag{6.B.4}$$

Combining (B.3) and (B.4) with (B.2) yields the required density:

$$h(t,s,|p,w) = \begin{array}{r} \frac{f(t|s,w)f(s|w)}{\int_0^\infty F(p|\tilde{s},w)f(\tilde{s}|w)ds} \\ 0 < s < \infty \\ p < t < \infty \end{array} \tag{6.B.5}$$

Chapter 7

Conclusions

The essays in this thesis deal with different aspects of microeconometric modelling of the labour market. Chapters 2, 3 and 4 concentrate on static models of labour supply, whereas in chapters 5 and 6, labour market transitions are modelled.

Chapter 2 presents three method of simulated scores (MSS) estimators that extend the method of simulated moments (MSM) estimator by McFadden (1989) to models with mixed discrete-continuous variables. The estimators are obtained by solving a system of moment equations that are derived from the score vector of the likelihood function. The moment equations are written such that simulators for probabilities and derivatives of probabilities enter the equations linearly. Consequently, simulation errors, which are defined as the difference between the simulators and the probabilities that are simulated, are averaged out over individuals, which implies that the parameter vector that solves the moment conditions is a consistent estimator, even if a fixed, finite number of replications is used in the calculation of the simulators. This is a property which is not shared by the simulated maximum likelihood estimator (SML), which is obtained by inserting simulated probabilities into the likelihood function. For consistency of the latter method, both sample size and number of drawings should tend to infinity. Lerman and Manski (1981) applied this method while using a frequency simulator. They found that a large number of replications was required to reduce the bias, that is introduced by using SML, to acceptable limits. Recently, interest in the SML method has been renewed by the application of SML in the context of multidimensional Probit and Tobit models. Börsch-Supan and Hajivassiliou (1993) propose the use of smooth simulators for the application of SML, instead of using frequency simulators. They simulate multidimensional normal probability integrals by means of the smooth recursive conditioning algorithm (SRC). They show that the application of SML to a multidimensional Probit model, simulating the probabilities by means of SRC, already yields small biases for ten drawings, whereas with twenty replications the bias has been reduced so far that a further increase in replications does not result in any further improvement. In the implementation of SRC the convenient properties of the normal distribution are exploited. Moreover, the applicability of the method depends on the explicit availability of integration bounds. The reason for application of frequency simulators, however, is the absence of explicit expressions for the bounds of integration. The motivation for the construction of MSS estimators in chapter 2 is to provide methods of estimation by simulation that are able to cope with frequency simulators as well as general non-normal probability distributions.

The MSS methods are applied to a simple neo-classical model of labour supply in which the number of working hours and the wage rate are modelled simultaneously. The simplicity of the model allows for comparison of the MSS methods with the maximum likelihood estimator (ML), as well as with SML. Both Monte Carlo and empirical results are presented. The Monte Carlo study shows that the MSS estimators perform satisfactorily, even if a small number of drawings to simulate the response probabilities, is used. The Monte Carlo study also shows that the MSS methods outperform smooth simulated maximum likelihood if a small number of drawings is used in the simulation of the response probabilities. In the empirical application, substantial differences appear in the estimation results obtained with different estimation methods. These differences are also revealed in the computed wage elasticities of hours and participation. The differences are attributed to the simplicity of the model, which ignores issues like fixed cost of work, taxation and demand side restrictions.

In chapter 3, one of the MSS estimators developed in chapter 2 is applied to a more sophisticated neo-classical labour supply model, which incorporates a piecewise linear and possibly non-convex budget set. Coherency conditions on parameters and supports of error distributions are imposed for all observations. Three stochastic error terms are introduced to present respectively optimization and reporting errors, stochastic preferences and heterogeneity in wages. The performance of the MSS estimator is compared with some more conventional estimation methods that are proposed in the literature. Monte Carlo, as well as empirical results, obtained with the various methods of estimation, are presented. The results show large variation of outcomes across estimation methods. The ordering of estimated wage and participation elasticities, obtained by applying the methods of estimation to real data, is by and large the same as for the Monte Carlo data. The results seem to show the importance of a correct (utility consistent) treatment of the stochastic structure of the model. The performance of the MSS estimator opens possibilities for further sophistication of the model. In particular, one may be interested in the introduction of fixed cost of work, like cost of childcare, travelling expenses etc., which generates an entry barrier. The MSS method seems to be well suited for handling these kind of additional complications.

Chapter 4 presents a model of labour supply in which individuals are faced by demand side restrictions. It is based on models by Dickens and Lundberg (1985), Tummers and Woittiez (1991) and Van Soest, Woittiez and Kapteyn (1990). In these papers the individual is assumed to be offered a random number (possibly zero) of job offers. The job offers carry the same gross wage rate, but may have different numbers of weekly working hours. The job yielding the highest level of utility is selected and its level of utility is compared with the utility level of not working, after which the participation decision is made. The probability of obtaining a job offer is assumed to be the same for every individual, irrespective of age, level of education, etc. Moreover, the gross wage rate is the same for every job offer, which differs from job search theory (Mortensen (1986)), where the wage rate differs across job offers and hours of work usually are not taken into consideration. The model that is presented in chapter 4 extends the Dickens and Lundberg (1985) model by allowing wages to vary across job offers. Moreover, the job offer probability depends on individual characteristics. A likelihood ratio test rejects the hypothesis that the probability of receiving a job offer is independent of individual characteristics. The estimates of the utility parameters in the extended model, in which the distribution of the number of job offers depends on individual characteristics, are estimated very imprecisely. As a result, preferences are largely dominated by the random component in the utility function. Due to the assumption that all job offers arrive at a given point in time, while the actual number of job offers is not observed, the data on the endogenous variables, hours and wages, apparently contain too little information for obtaining precise estimates of both offer probabilities and utility parameters. To improve on the performance of the model, it is proposed to drop the assumption that job offers arrive at a given point in time and replace it by introducing a sequential arrival scheme of job offers. Data on unemployment duration can then be used as an additional source of information in the estimation of the model parameters. This calls for further extension of the model with elements of job search theory and this extension is the subject of chapter 5.

Chapter 5 presents a model of sequential job search, in which a job consists of a wage component and an hours component. Two variants of the model are presented. The first assumes that once a job offer with a specific wage rate is obtained, the individual is allowed to determine the number of working hours by himself. The second variant assumes that a job offer is a package consisting of a wage rate and a number of working hours. For the first model it is derived that there exists a reservation wage strategy. The individual accepts job offers with a wage rate that exceeds the reservation wage rate. The reservation wage rate is determined by the benefit level, the job offer arrival rate, the wage distribution, as well as by optimal labour supply, which depends on the determinants of the utility function. After the acceptance of a job, labour supply is determined in a way that is comparable to the neo-classical labour supply models in chapters 2 and 3. The difference with the neo-classical model is that in the latter model there is no separation between the participation decision and the labour supply decision. In the present model, this separation is introduced in a structural way, as labour supply depends on preferences only, whereas participation depends on preferences as well as on the job offer arrival rate. The assumptions underlying the second model imply that a reservation utility strategy is followed. Specific attention is paid to the stochastic specification of the model and various sources of randomness are introduced. Data on unemployment duration and after unemployment spell job characteristics are used to estimate the model. This model is the proposed extension of the model in chapter 4. Some plots, based on residual analysis, reveal that the second model with hours constraints outperforms the model without hours constraints. The plots also show the presence of positive duration dependence of the hazard rate. Because of the complexity of the models, we have abstained from taking into account the tax system, as was done in chapter 3. Although it is in principle possible to incorporate the tax system, the practical implementation would become more burdensome, both in terms of tractability of quantities like the reservation wage, and in terms of required computing time.

Chapter 6 presents a structural model of job search in which the job offer arrival rate is endogenous. Individuals can influence their job offer arrival rate by varying their search effort. The optimal amount of search effort is determined on basis of a comparison of marginal returns of search and marginal cost of search, under the restriction that search effort is non-negative. In case of a corner solution, it is optimal not to search. The model studies the search behaviour of both unemployed and employed individuals. We allow for differences in search conditions for unemployed and employed individuals, and it turns out that these differences play an important role in the determination of the reservation wage rate of unemployed persons. In taking the job acceptance decision, unemployed individuals take these differences in search conditions into account. Search effort requires an operational definition and several indicators of search effort are available in the dataset. In the model, search effort is allowed to be a vector, of which each component is related to a specific indicator appearing in the dataset. The presence of four indicators of search intensity in the dataset requires the evaluation of up to four-dimensional probability integrals in calculating the maximum likelihood estimator. A multivariate normal distribution is specified for the underlying error structure of the four search effort indicators. Because of the assumed normality, together with the presence of explicitly defined integration regions, the present model fits perfectly in the smooth simulated maximum likelihood (SSML) framework of Börsch-Supan and Hajivassiliou (1993), and we use the smooth recursive conditioning algorithm to simulate the multidimensional normal probability integrals. Estimation results show that there is a positive effect of expected returns of search on optimal search intensity, which is in accordance with the predictions of the economic model. For both employed and unemployed individuals, the individual's attitude towards search turns out to have substantial influence on the probability of getting a job. For employed individuals, age patterns are a significant determinant of the cost of search, whereas for unemployed individuals, variables associated with household composition significantly influence cost of search. The emphasis in this study is on detecting the influence of search intensity on unemployment duration and job tenure, as well as on the determination of cost of search characteristics. For these purposes, the present model specification, based on the model of Mortensen (1986), is amply sufficient. However, one may want to proceed and model jointly the decision on time spent on labour supply and time spent on search, as in the model of Burdett and Mortensen (1978). It is clear that for an empirical implementation of the latter model, which would amount to a combination of the chapters 5 and 6 of this thesis, more detailed information about search intensity is required. One would need data on search intensity in terms of the weekly amount of hours spent on search activities.

Although there are several links between the models appearing in the various chapters, a higher degree of integration of the models may be desirable. In each of the chapters, one specific issue of microeconometric modelling of labour market behaviour is highlighted and in each of the chapters the importance of incorporating this specific issue is stressed. Ideally, one would like to build these several issues into one model. In doing this, we will likely be hampered by the complexity that will arise in the implementation of such a model: Even with rather simple specifications of error distributions and utiltity functions models may become intractable, either because we cannot obtain explicit functional relationships, or because numerical methods of evaluation require substantial amounts of computing time.

Throughout this thesis, simulation estimators have been proven to be useful tools in the empirical implementation of structural models of the labour market. Two basic simulation estimators have been used. The first estimator is the method of simulated scores (MSS). The attractiveness of this estimator is that it is consistent for a fixed number of drawings. It is suitable for the use of both frequency simulators and smooth simulators. The second estimator is simulated maximum likelihood (SML), which is not consistent for a small, fixed number of drawings. Implementation of this method while using frequency simulators has proven not to be very succesful. However, recent work shows that SML with smooth multivariate normal probability simulators, performs well in practice, even with a limited number of replications like ten. Frequency simulators are useful for models with complicated, implicitly defined choice sets. A disadvantage of the frequency simulator is its discontinuity, which makes standard optimization routines, based on gradient methods, inappropriate. Non-gradient methods are usually costly in terms of computing time. The SRC simulator turns out to be a convenient simulator for the simulation of multivariate normal probability integrals. Whether the SRC simulator will be a practical solution for every arbitrary multivariate distribution and whether it still performs well in combination with SML in the case of non-normal distributions, is scope for further research.

References

Abowd, J. and O. Ashenfelter, 1981, Anticipated unemployment, temporary layoffs, and compensating wage differentials, in S. Rosen (ed.), *Studies in labor markets*, University of Chicago Press, 141-170.

Andrews, D.W.K, 1988, Chi-square Diagnostic Tests for Econometric Models, *Journal of Econometrics* 37, pp. 135-156

Berg, G.J. van den, 1990a, *Structural dynamic analysis of individual labour market behaviour*, Ph.D. thesis.

Berg, G.J. van den, 1990b, Nonstationarity in job search theory, *Review of Economic Studies* 57, 225-277.

Berg, G.J. van den, 1990c, Search behaviour, transitions to nonparticipation and the duration of unemployment, *Economic Journal*, Vol. 100, pp. 842-865.

Berg, G.J. van den, 1992, A structural dynamic analysis of job turnover and the costs associated with moving to another job, *Economic Journal*, Vol. 102, No. 414, pp. 1116-1133.

Blau, D.M, 1991, Search for nonwage job characteristics: a test of the reservation wage hypothesis, *Journal of Labor Economics*, vol. 9, no. 2, 186-205.

Bloemen, H.G. and A. Kapteyn, 1993a, The joint estimation of a non-linear labour supply function and a wage equation using simulated response probabilities, *Annales d'Economie et de Statistiques*, Nr. 29, pp. 175-205.

Bloemen, H.G. and A. Kapteyn, 1993b, The estimation of utility consistent labor supply models by means of simulated scores, CentER Discussion Paper No. 9367.

Blomquist, N, 1983, The effect of income taxation on the labour supply of married men in Sweden, *Journal of Public Economics*, Vol. 22, pp. 169-197.

Blundell, R.W, J. Ham and C. Meghir, 1987, Unemployment and Female Labour Supply, *Economic Journal*, 97, pp. 44-64

Blundell, R. and C. Meghir, 1987, Bivariate alternatives to the Tobit model, *Journal of Econometrics*, Vol. 34, pp. 179-200.

Börsch-Supan, A. and V. Hajivassiliou, 1993, Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models, *Journal of Econometrics*, Vol. 58, nr. 3, pp. 347-368.

Burdett, K. and D.T. Mortensen, 1978, Labour supply under uncertainty, *Research in labour economics*, Vol. 2, 109-157.

Cox, D.R. and D. Oakes, 1984, *Analysis of survival data*, Chapman and Hall, New York.

Devroye, L, 1986, *Non-uniform random variate generation*, Springer Verlag, New York.

Dickens, W. and S. Lundberg, 1985, Hours restrictions and labor supply, NBER working paper.

Flinn, C. and J. Heckman, 1982a, New methods for analyzing structural models of labor force dynamics, *Journal of Econometrics* 18, 115-168.

Flinn, C. and J. Heckman, 1982b, Models for the analysis of labor force dynamics, *Advances in Econometrics* vol. 1, 35-95.

Gouriéroux, C. and A. Monfort, 1989, Simulation Based Inference in Models with Heterogeneity, mimeo, report 8905, INSEE, Paris

Gouriéroux, C., J. Laffont and A. Montfort, 1980, Coherency conditions in simultaneous linear equation models with endogenous switching regimes, *Econometrica* 48, pp. 675-695.

Hajivassiliou, V, 1989, Simulated Scores for Estimating LDV Models: A Presentation and Comparative Evaluation, mimeo, Cowles Foundation

Hajivassiliou, V, 1992, Simulation Estimation Methods for Limited Dependent Variable Models, in: *Handbook of Statistics*, Vol. 11 (Econometrics), (forthcoming), North-Holland.

Hammersley, J.M. and D.C. Handscomb, 1964, *Monte Carlo Methods*, London: Methuen.

Hausman, J, 1979, The econometrics of labor supply on convex budget sets, *Economics letters* 3, pp. 171-174.

Hausman, J, 1980, The Effect of Wages, Taxes and Fixed Costs on Women's Labor Force Participation, *Journal of Public Economics* 14 pp. 161-194

Hausman, J, 1981, Labor supply, *How Taxes Affect Economic Behavior*, ed. H. J. Aaron and J. A. Pechman, The Brookings Inst., Washington DC, pp. 27-83

Hausman, J, 1985, The econometrics of nonlinear budget sets, *Econometrica* 53, pp. 1255-1282.

Hausman, J. and P. Ruud, 1984, Family labor supply with taxes, *American Economic Review*, Vol. 74, pp. 242-248

Heckman, J, 1974, Shadow Prices, Market Wages, and Labor Supply, *Econometrica*, vol. 42, no. 4, pp. 679-694

Kapteyn, A., P. Kooreman and A. van Soest, 1990, Quantity Rationing and Concavity in a Flexible Household Labor Supply Model, *The Review of Economics and Statistics*, Vol. 72, pp. 55-62

Kapteyn, A. and I. Woittiez, 1989, Preference interdependence and habit formation in family labour supply, in: *Microeconometrics: Surveys and applications*, ed. by F. Laisney et al., Clevedon, Avon, Tieto Ltd., Chapter 10.

Kiefer, N.M, 1984, Employment contracts, job search theory, and an empirical model of labor turnover, working paper, Cornell University.

Kiefer, N.M, 1987, A proposition and an example in the theory of job search with hours constraints, *Journal of Labor Economics*, vol. 5, 211-220.

Kiefer, N.M, 1988, Economic duration data and hazard functions, *Journal of Economic Literature*, vol. XXVI, 646-679.

Kloek, T. and H.K. van Dijk, 1978, Bayesian estimates of equation system parameters: an application of integration by Monte Carlo, *Econometrica*, vol. 46, no. 1, 1-19.

Kooreman, P. and A. Kapteyn, 1986, Estimation of rationed and unrationed household labor supply functions using flexible functional forms, *The Economic Journal*, Vol. 96, pp. 398-412.

Lancaster, T, 1990, *The econometric analysis of transition data*, Cambridge University Press, Cambridge.

Lerman, C. and C. Manski, 1981, On the Use of Simulated Frequencies to Approximate Choice Probabilities, in: *Structural Analysis of Discrete Data with Economic Applications*, ed. C. Manski and D. McFadden, MIT Press, Cambridge, pp. 305-319

Lindeboom, M. and J. Theeuwes, 1992, Search, benefits and entitlement, Research Memorandum, Leiden University.

MaCurdy, T, D. Green and H. Paarsch, 1990, Assessing empirical approaches for analyzing taxes and labor supply, *Journal of Human Resources*, Vol. 24, nr. 3, pp.415-490.

McFadden, D, 1989, A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, *Econometrica*, Vol. 57, pp. 995-1026

Mc.Kenna, C.J, 1985, *Uncertainty and the labour market*, Wheatsheaf Books, Brighton.

Moffitt, R, 1982, The Tobit model, hours of work and institutional constraints, *The Review of Economics and Statistics*, Vol. LXIV, nr. 3, 510-515.

Moffit, R, 1984, The Estimation of a Joint Wage-hours Labour Supply Model, *Journal of Labour Economics* 2, pp. 550-566

Moffit, R, 1986, The econometrics of piecewise-linear budget constraints, *Journal of Business and Economic Statistics* 4, pp. 317-328.

Mortensen, D.T, 1986, Job search and labour market analysis, in: O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics*, Vol. II, Elsevier, Amsterdam.

Narendranathan, W. and S.J. Nickell, 1985, Modelling the process of job search, *Journal of Econometrics* 28, 29-49.

Nelder, J.A. and R. Mead, 1965, A simplex method for function minimization, *Computer Journal* Vol. 7, pp. 308-313.

Pakes, A. and D. Pollard, 1989, Simulation and the Asymptotics of Optimization Estimators, *Econometrica*, Vol. 57, pp. 1027-1057

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, 1986, *Numerical Recipes*, Cambridge University Press, Cambridge

Rettore, E, 1990, Institutional constraints on working week length and female labour supply, working paper, University of Padova.

Ridder, G, 1984, The distribution of single-spell duration data, in: G.R. Neumann and N. Westergård-Nielsen, eds., *Studies in labor market analysis*, Springer Verlag, Berlin.

Ridder, G, 1987, *Life cycle patterns in labor market experience*, Ph.D. thesis, Amsterdam.

Stroud, A.H. and D. Secrest, 1966, *Guassian quadrature formulas*, Prentice-Hall

Tummers and Woittiez, 1991, A simultaneous wage and labor supply model with hours restrictions, *Journal of Human Resources*, vol 26, nr. 3, 213-226.

Van Soest, A., P. Kooreman and A. Kapteyn, 1993, Coherency and Regularity of Demand Systems with Equality and Inequality Constraints, *Journal of Econometrics*, Vol. 57, nr. 1-3, pp. 161.

Van Soest, A.H.O, I. Woittiez and A. Kapteyn, 1990, Labour supply, income taxes and hours restrictions in the Netherlands, *Journal of Human Resources*, vol. 25, nr. 3, 517-558.

Yoon, B.J, 1981, A model of unemployment duration with variable search intensity, *Review of Economics and Statistics*, Vol. LXII, No. 4, pp. 599-609.

Samenvatting

In dit proefschrift worden structurele modellen van arbeidsmarktgedrag van individuën geformuleerd en geschat. Alle modellen zijn geformuleerd op basis van economische gedragsveronderstellingen, zoals nutsmaximalisatie onder inkomensrestricties. In de hoofdstukken 2, 3 en 4 komen statische arbeidsaanbodmodellen aan bod. De hoofdstukken 5 en 6 bespreken modellen waarin transities tussen verschillende arbeidsmarkttoestanden mogelijk zijn. Deze modellen zijn gebaseerd op de zoektheorie.

Verscheidene bronnen van onzekerheid worden in de modellen opgenomen, waaronder onzekerheid ten aanzien van individuele preferenties en meet- en optimalisatiefouten. De parameters van de modellen worden geschat met behulp van econometrische technieken. Hierbij wordt veelvuldig gebruik gemaakt van de zogenaamde simulatieschatters. Simulatieschattingsmethoden zijn gebaseerd op momentenmethoden en maximale aannemelijkheidsmethoden, maar zij verschillen van deze methoden doordat integralen, die verdelingsmomenten representeren, vervangen worden door sommen die berekend worden met behulp van kansgetallen welke uit een kansverdeling getrokken worden. Er zijn verschillende redenen waarom men simulatoren voor kansen zou willen gebruiken. Ten eerste kan het economische model dermate ingewikkeld zijn dat expliciete uitdrukkingen voor het integratiegebied van de kansintegralen niet bestaan. Dit integratiegebied komt overeen met een bepaalde keuzeverzameling welke door het model gedefiniëerd wordt. Doorgaans is het echter wel mogelijk om kansgetallen te trekken uit de in het model gespecificeerde verdeling, om vervolgens na te gaan of men zich wel of niet in het betreffende keuzegebied bevindt. Door deze procedure herhaaldelijk en voor verschillende individuën uit te voeren kan een frequentie bepaald worden welke als benadering voor de kans gebruikt kan worden. Ten tweede kan men modellen tegenkomen waarin het aantal keuzemogelijkheden of het aantal kansvariabelen zo hoog is dat de dimensie van de te berekenen kansintegralen te hoog is om de schattingsprocedure met behulp van numerieke integratiemethoden binnen een praktisch aanvaardbare computertijdsperiode te berekenen. Twee simulatieschatters worden gebruikt: de methode van gesimuleerde scores en de methode van gesimuleerde maximale aannemelijkheid.

Hoofdstuk 1 bevat een inleidend overzicht van het proefschrift.

In hoofdstuk 2 worden verscheidene simulatieschatters geïntroduceerd welke gebaseerd zijn op het simuleren van de scorevector van de aannemelijkheidsfunctie. Een voordeel van de score simulatieschatters is dat zij consistent zijn, zelfs als het aantal trekkingen dat gebruikt wordt om kansen te simuleren eindig is. Dit in tegenstelling tot de methode van gesimuleerde maximale aannemelijkheid waarbij de gesimuleerde kansen worden ingevuld in de oorspronkelijke aannemelijkheidsfunctie. Voor deze laatste methode geldt dat het aantal trekkingen naar oneindig moet gaan voor het verkrijgen van consistente schatters.

De simulatieschatters worden toegepast op een eenvoudig neo-klassiek arbeidsaanbodmodel. Om de werking van de schatters te bestuderen worden zij toegepast op Monte Carlo dataverzamelingen. De scoreschatters blijken bevredigend te functioneren, zelfs indien slechts een beperkt aantal trekkingen gebruikt wordt om de kansen te simuleren. De methode van gesimuleerde maximale aannemelijkheid blijkt meer trekkingen nodig te hebben om tot goede resultaten te komen. Vervolgens worden de schattingsmethoden toegepast op een steekproef van getrouwde vrouwen in Nederland in 1985. De verschillende methoden blijken resultaten te genereren die behoorlijk van elkaar verschillen. Dit komt met name tot uitdrukking in de schattingen van de loonelasticiteiten van arbeidsaanbod en participatie. De verschillen worden geweten aan de eenvoud van het model, waarin geen rekening gehouden is met zaken als vaste kosten van arbeid, het belastingstelsel en restricties aan de vraagzijde van de arbeidsmarkt.

In hoofdstuk 3 wordt een uitgebreider arbeidsaanbodmodel geformuleerd. Hierin wordt tevens het belastingstelsel opgenomen. Als gevolg hiervan is de inkomensrestrictie stuksgewijs lineair en mogelijk niet-convex. Er wordt een coherentierestrictie opgelegd die moet bewerkstelligen dat de endogene variabelen eenduidig uit het model oplosbaar zijn, en het verband tussen deze coherentierestrictie en de regulariteitsvoorwaarden voor de nutsfunctie wordt benadrukt. Het model bevat meerdere toevalsvariabelen. Er is een toevalsvariabele geïntroduceerd in de preferentiestructuur, meet- en optimalisatiefouten in uren komen voor en er wordt rekening gehouden met heterogeniteit in lonen. Het model is dusdanig gecompliceerd dat geen expliciete uitdrukkingen voor het integratiegebied van de kansintegralen voor handen zijn. Daarom worden de responsiekansen gesimuleerd met behulp van frequentiesimulatoren. De in hoofdstuk 2 beschreven scoreschatters zijn uitermate geschikt voor de toepassing van frequentiesimulatoren. Verscheidene in de literatuur bestaande schattingsmethoden, welke gebaseerd zijn op benaderingen of vereenvoudigingen van het oorspronkelijke model worden vergeleken met de score simulatieschatters. Het betreft hier het verwaarlozen van het toevalselement in de preferenties en het schatten met instrumentele variabelen, waarbij de inkomensrestrictie wordt gelineariseerd. De schattingsmethoden worden met elkaar vergeleken in een Monte Carlo studie, waarbij het met name opvalt dat er grote verschillen voorkomen in de schattingen van de loonelasticiteiten. De score simulatieschatter blijkt parameterschattingen te genereren die aanmerkelijk dichter bij de ware parameterwaarden liggen dan de benaderingsmethoden. De instrumentele variabelen schatter maakt geen gebruik van de coherentierestrictie. Bij controle achteraf blijkt dat voor de instrumentele variabelen schattingen niet alle individuën aan deze coherentierestrictie voldoen, terwijl voor de overige methoden, waarbij de coherentierestrictie gedurende de schattingsprocedure wordt opgelegd, deze restrictie niet bindend blijkt te zijn. De resultaten van de Monte Carlo studie benadrukken het belang van het in acht nemen van de volledige structuur van het model, in plaats van het gebruik van benaderingen. Hierna worden de methoden toegepast op een empirische dataset. Ook hier komen weer verschillen in de loonelasticiteiten tussen de diverse methoden naar voren.

Hoofdstuk 4 beschrijft een arbeidsaanbodmodel waarin individuën geconfronteerd worden met restricties aan de vraagzijde van de arbeidsmarkt. Er wordt verondersteld dat iemand op een gegeven tijdstip een door toeval bepaalde hoeveelheid banen krijgt aangeboden, waarbij een baan gekarakteriseerd wordt door een loonvoet en een vast aantal arbeidsuren per week. De baan die het hoogste nutsniveau oplevert wordt geselecteerd en vervolgens wordt dit nutsniveau vergeleken met het nut van niet werken, waarna de acceptatiebeslissing wordt genomen. De kansverdeling die het aantal aangeboden banen beschrijft is afhankelijk gemaakt van individuele eigenschappen. Een toets verwerpt de hypothese dat deze afhankelijkheid niet geldt.

Hoofdstuk 5 breidt dit model uit met elementen uit de zoektheorie. Er wordt uitgegaan van werkloze individuën die op zoek zijn naar een baan, waarbij een baan wordt gekenmerkt door twee eigenschappen, namelijk de loonvoet en het aantal arbeidsuren per week. Twee modellen worden in beschouwing genomen. In het eerste model wordt aangenomen dat als iemand een baan met een gegeven loonvoet aangeboden heeft gekregen, hij vervolgens zelf mag bepalen hoeveel uren per week hij wil werken. In dit geval kan worden aangetoond dat er een reserveringsloon bestaat. Alleen een aanbod met een loon dat hoger is dan het reserveringsloon wordt geaccepteerd. In het tweede model bepaalt de werkgever zowel het loon als het aantal arbeidsuren. Nu bestaat er een reserveringsnutsniveau. Alleen loon-uren combinaties die een hoger nutsniveau opleveren dan het reserveringsnutsniveau worden geaccepteerd. De modellen worden geschat met gegevens over geaccepteerde uren en lonen en gegevens over de werkloosheidsduur. Een vergelijking van beide modellen vindt plaats aan de hand van residuele analyse. Hieruit blijkt dat het tweede model de data beter beschrijft dan het eerste.

In hoofdstuk 6 wordt een zoekmodel geformuleerd waarin individuën zelf de kans op een baanaanbod kunnen beïnvloeden door hun zoekintensiteit te variëren. Een hogere zoekintensiteit (bijvoorbeeld aantal sollicitaties) leidt tot een hogere kans op een baanaanbod, maar tevens tot hogere zoekkosten. Marginale opbrengst en marginale kosten van zoeken worden met elkaar vergeleken om de optimale zoekintensiteit te bepalen. Hierbij wordt onderscheid gemaakt tussen zij die werkloos zijn en zij die zoeken terwijl zij een baan hebben. De dataset bevat verschillende indicatoren die informatie bevatten over de zoekintensiteit. Er is een indicator die informatie geeft over de houding van zoekenden, i.e. zoekt men serieus, of zoekt men minder serieus, er is een variabele die aangeeft of men de personeelsadvertenties bijhoudt en er is een variabele die aangeeft of men al dan niet staat ingeschreven bij het arbeidsbureau. Verder is er informatie over sollicitatieaantallen. Met name de houdingsvariabele blijkt van invloed te zijn op de aanbodkans. Opmerkelijk is dat voor werkenden het aantal sollicitaties weinig effect lijkt te hebben. Voor werklozen is dat wel het geval. Voor werklozen is ingeschreven staan bij het arbeidsbureau het minst effectieve zoekinstrument. Het economische model voorspelt dat een hogere marginale opbrengst van zoeken leidt tot een hogere zoekintensiteit. De empirische resultaten onderschrijven dit. De zoekkosten van werkenden lijken te variëren over de levenscyclus, terwijl voor werklozen de gezinssituatie eveneens een rol speelt.

Hoofdstuk 7 geeft een overzicht van de voorgaande hoofdstukken, waarbij enige concluderende opmerkingen geplaatst worden.